SPARC Data Initiative: A comparison of ozone climatologies from international satellite limb sounders

S. Tegtmeier,[1] M. I. Hegglin,[2] J. Anderson,[3] A. Bourassa,[4] S. Brohede,[5,18] D. Degenstein,[4] L. Froidevaux,[6] R. Fuller,[6] B. Funke,[7] J. Gille,[8] A. Jones,[9] Y. Kasai,[10] K. Krüger,[1,19] E. Kyrölä,[11] G. Lingenfelser,[12] J. Lumpe,[13] B. Nardi,[14] J. Neu,[6] D. Pendlebury,[9] E. Remsberg,[12] A. Rozanov,[15] L. Smith,[14] M. Toohey,[1] J. Urban,[5] T. von Clarmann,[16] K. A. Walker,[9] and R. H. J. Wang[17]

Received 20 March 2013; revised 2 October 2013; accepted 3 October 2013; published 8 November 2013.

[1] A comprehensive quality assessment of the ozone products from 18 limb-viewing satellite instruments is provided by means of a detailed intercomparison. The ozone climatologies in form of monthly zonal mean time series covering the upper troposphere to lower mesosphere are obtained from LIMS, SAGE I/II/III, UARS-MLS, HALOE, POAM II/III, SMR, OSIRIS, MIPAS, GOMOS, SCIAMACHY, ACE-FTS, ACE-MAESTRO, Aura-MLS, HIRDLS, and SMILES within 1978–2010. The intercomparisons focus on mean biases of annual zonal mean fields, interannual variability, and seasonal cycles. Additionally, the physical consistency of the data is tested through diagnostics of the quasi-biennial oscillation and Antarctic ozone hole. The comprehensive evaluations reveal that the uncertainty in our knowledge of the atmospheric ozone mean state is smallest in the tropical and midlatitude middle stratosphere with a 1σ multi-instrument spread of less than ±5%. While the overall agreement among the climatological data sets is very good for large parts of the stratosphere, individual discrepancies have been identified, including unrealistic month-to-month fluctuations, large biases in particular atmospheric regions, or inconsistencies in the seasonal cycle. Notable differences between the data sets exist in the tropical lower stratosphere (with a spread of ±30%) and at high latitudes (±15%). In particular, large relative differences are identified in the Antarctic during the time of the ozone hole, with a spread between the monthly zonal mean fields of ±50%. The evaluations provide guidance on what data sets are the most reliable for applications such as studies of ozone variability, model-measurement comparisons, detection of long-term trends, and data-merging activities.

Citation: Tegtmeier, S., et al. (2013), SPARC Data Initiative: A comparison of ozone climatologies from international satellite limb sounders, *J. Geophys. Res. Atmos.*, *118*, 12,229–12,247, doi:10.1002/2013JD019877.

1. Introduction

[2] Stratospheric ozone is one of the most important trace gases in the atmosphere due to its absorption of biologically harmful ultraviolet radiation and its role in determining the temperature structure of the atmosphere. The depletion of stratospheric ozone as a result of anthropogenic emissions of halogens is expected to decrease and reverse [*Austin and Butchart*, 2003; *SPARC CCMVal*, 2010; *WMO*, 2011] due to the phaseout of ozone-depleting substances specified by the Montreal Protocol and its subsequent amendments. Detection and attribution of the expected ozone recovery in a future changing climate [e.g., *Newman et al.*, 2006;

[1]GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany.
[2]University of Reading, Reading, UK.
[3]Hampton University, Hampton, Virginia, USA.
[4]University of Saskatchewan, Saskatoon, Saskatchewan, Canada.
[5]Chalmers University, Göteborg, Sweden.
[6]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA.

Corresponding author: S. Tegtmeier, GEOMAR Helmholtz Centre for Ocean Research Kiel, Düsternbrooker Weg 20, D-24105 Kiel, Germany. (stegtmeier@geomar.de)

2169-897X/13/10.1002/2013JD019877

[7]Instituto de Astrofísica de Andalucía, Granada, Spain.
[8]University of Colorado, Boulder, Colorado, USA.
[9]University of Toronto, Toronto, Ontario, Canada.
[10]NICT, Koganei, Japan.
[11]Finnish Meteorological Institute, Helsinki, Finland.
[12]NASA Langley Research Center, Hampton, Virginia, USA.
[13]Computational Physics, Inc., Boulder, Colorado, USA.
[14]National Center for Atmospheric Research, Boulder, Colorado, USA.
[15]Institute of Environmental Physics (IUP), University of Bremen, Bremen, Germany.
[16]Karlsruhe Institute of Technology, Karlsruhe, Germany.
[17]Georgia Institute of Technology, Atlanta, Georgia, USA.
[18]Now at FluxSense AB, Göteborg, Sweden.
[19]Now at University of Oslo, Oslo, Norway.

Waugh et al., 2009] with increasing greenhouse gases and a modified residual circulation require a comprehensive understanding of short- and long-term ozone changes and their vertical, latitudinal, and seasonal dependence. Such knowledge can only be derived from vertically resolved, high-quality, global, long-term observational data sets.

[3] Among the different measurement systems that provide ozone observations with high vertical resolution, satellite instruments are the most suitable option for creating spatially continuous data sets. A large number of limb-viewing satellite instruments have been measuring stratospheric ozone over the past three decades, providing a wealth of vertically resolved ozone observations. The individual satellite data sets vary in terms of measurement method, geographical and temporal coverage, resolution, and retrieval algorithm and can therefore deviate from each other. It is often difficult for a user to determine which satellite instrument provides the most reliable or useful data set for a particular application. While ozone records are widely used for the validation of transport and chemistry in numerical models [e.g., *SPARC CCMVal*, 2010], such comparisons of observations and model output can become less meaningful without the detailed knowledge of the quality and details of the observations. Furthermore, realistic ozone data sets are important input fields for global climate models that do not include interactive chemistry in order to reproduce climate responses such as Southern Hemisphere (SH) tropospheric circulation and surface temperature changes [*Gillett and Thompson*, 2003; *Son et al.*, 2009]. Another focus of current research is the detection of ozone profile trends [*WMO*, 2011 and references therein], in particular for the time period after 2005. The SAGE II ozone data set is considered to be the most reliable long-term satellite data source for the detection and quantification of ozone changes in the lower stratosphere [e.g., *Randel and Wu*, 2007]. However, SAGE II covers the time period between 1984 and 2005, and while many newer satellite instruments have provided vertical ozone distribution since 2000, a thorough assessment of the newer measurements with each other and the older data sets is critical in order to create a merged data set that can extend ozone trend analysis beyond the lifetime of the SAGE II instrument.

[4] A large number of studies have focused on the validation of individual satellite ozone data sets by means of coincident measurement comparisons [e.g., *Randall et al.*, 2003; *Jiang et al.*, 2007; *Steck et al.*, 2007; *Froidevaux et al.*, 2008; *Livesey et al.*, 2008; *Nardi et al.*, 2008; *Dupuy et al.*, 2009; *Kyrölä et al.*, 2010; *Mieruch et al.*, 2012]. Additionally, there are ozone data merging activities that focus either on the European satellite missions or on the USA and Canadian satellite missions. These activities include detailed intercomparisons of several data records [e.g., *Jones et al.*, 2009]; however, no single comparison of all available ozone data sets from international limb sounders has been available so far.

[5] The first comprehensive intercomparison of ozone data sets available from limb-viewing satellite instruments was performed as part of the Stratosphere-troposphere Processes And their Role in Climate (SPARC) Data Initiative (M. I. Hegglin et al., SPARC Data Initiative: Comparison of trace gas and aerosol climatologies from international satellite limb sounders, manuscript in preparation, 2013) and is presented in this paper. The comparisons will provide basic information on quality and consistency of the various ozone products and will serve as a guide for their use in empirical studies of climate and variability and in model-measurement comparisons. Ozone observations available from 1978 until the end of 2010 from 18 international satellite instruments are included in the comparison, and the spread in the climatologies is used to provide an estimate of the overall systematic uncertainty in our knowledge of the mean ozone state. The individual monthly zonal mean time series are compared in terms of their zonal mean climatologies, seasonal evolution (section 3), and interannual variability (section 4). Additionally, the physical consistency of the data sets is tested through diagnostics of the quasi-biennial oscillation (QBO) (section 4.1) and the Antarctic ozone hole (section 5).

2. Data and Methods

3. Satellite Instruments

[6] Ozone data products with a high vertical resolution from limb-viewing satellite instruments are the focus of this study. Limb-viewing sounders can be classified according to their measurement mode (emission, scattering, solar occultation, and stellar occultation) or the wavelength band in which they operate. The former classification determines sampling patterns and therefore horizontal coverage and resolution of the retrieved data sets. The instruments participating in the SPARC Data Initiative are given with their full instrument name, satellite platform, measurement mode, and wavelength category in Table 1. Detailed information on the individual instruments including their sampling patterns and retrieval techniques can be found in the (M. I Hegglin and S. Tegtmeier, SPARC Data Initiative report on the evaluation of trace gas and aerosol climatologies from satellite limb sounders, in preparation, 2013). Note that although the solar backscatter ultraviolet/2 instruments provide a long-term ozone record with excellent coverage and density, the data are nadir viewing only and not included here due to their limited vertical resolution [*McLinden et al.*, 2009].

3.1. Ozone Climatologies

[7] The ozone climatologies from the individual satellite instruments consist of monthly zonal mean time series calculated on the SPARC Data Initiative climatology grid using 5° latitude bins and 28 pressure levels. The monthly zonal mean ozone volume mixing ratio (VMR), the standard deviation, and the number of averaged data values are given for each month, latitude bin, and pressure level. Furthermore, the mean, minimum, and maximum local solar time, the average latitude, and the day of the month of the measurements used to produce the climatologies are provided. The time series of all variables will be publicly available from the SPARC Data Center as NetCDF files. An overview of the ozone measurement records between 1978 and 2010 from the satellite instruments participating in this study is given in Figure 1. Note that while the SPARC Data Initiative is an ongoing activity, the evaluations presented here focus on measurements until the end of 2010.

[8] The climatology construction comprises careful screening (according to recommendations given in relevant quality documents), latitude binning, and linear interpolation in log pressure to the pressure grid 300, 250, 200, 170, 150, 130, 115, 100, 90, 80, 70, 50, 30, 20, 15, 10, 7, 5, 3, 2, 1.5, 1,

Table 1. Full Instrument Name, Satellite Platform, Measurement Mode, and Wavelength Category of all Instruments Participating in the SPARC Data Initiative Given in Order of Satellite Launch Date[a]

Instrument	Full Name	Satellite Platform	Measurement Mode	Wavelength Category
LIMS	Limb Infrared Monitor of the Stratosphere	Nimbus 7	Emission	Mid-IR
SAGE I/II/III	Stratospheric Aerosol and Gas Experiment	AEM-2, ERBS, Meteor-3 M	Solar occultation	Near-IR VIS/UV
UARS-MLS	UARS-Microwave Limb Sounder	UARS	Emission	Microwave/Sub-mm
HALOE	The Halogen Occultation Experiment	UARS	Solar occultation	Mid-IR
POAM II/III	Polar Ozone and Aerosol Measurement	SPOT-3/4	Solar occultation	Near-IR, VIS/UV
SMR	Sub-Millimetre Radiometer	Odin	Emission	Microwave/Sub-mm
OSIRIS	Optical Spectrograph and Infrared Imager System	Odin	Scattering	VIS/UV
MIPAS	Michelson Interferometer for Passive Atmospheric Sounding	Envisat	Emission	Mid-IR
GOMOS	Global Ozone Monitoring by Occultation of Stars	Envisat	Stellar occultation	VIS/UV
SCIAMACHY	SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY	Envisat	Scattering	Near-IR VIS/UV
ACE-FTS	Atmospheric Chemistry Experiment (ACE) -Fourier Transform Spectrometer	SCISAT-1	Solar occultation	Mid-IR
ACE-MAESTRO	ACE-Measurement of Aerosol Extinction in the Stratosphere and Troposphere Retrieved by Occultation	SCISAT-1	Solar occultation	VIS/UV
Aura-MLS	Aura-Microwave Limb Sounder	Aura	Emission	Microwave/Sub-mm
HIRDLS	High Resolution Dynamics Limb Sounder	Aura	Emission	Mid-IR
SMILES	Superconducting Submillimeter-Wave Limb Emission Sounder	ISS	Emission	Microwave/Sub-mm

[a]IR, infrared; VIS, visible; UV, ultraviolet.

0.7, 0.5, 0.3, 0.2, 0.15, and 0.1 hPa. If necessary, a conversion from altitude to pressure levels is performed using retrieved temperature/pressure profiles or meteorological analysis (European Centre for Medium-Range Weather Forecasts (ECMWF), National Centers for Environmental Prediction, UK Met Office). The same information is used to convert data products retrieved as number densities to VMR where needed. A minimum of five measurements within each bin is required to calculate a monthly zonal mean, although mostly, many more measurements are available in each bin. Detailed information on the climatology construction including the screening process for each instrument can be found in the SPARC Data Initiative report.

[9] For each data set, the data version, time period, vertical range, and resolution, as well as relevant references, are given in Table 2. The SAGE I climatology includes altitude corrections [*Wang et al*., 1996]. The SAGE III climatology does not include a separate retrieval for mesospheric ozone; therefore, it should be used with care. Note that UARS-MLS data are not as good for trend studies after June 1997 as a result of sparser data and missing MLS temperature retrievals. The MIPAS climatologies for 2002–2004, when MIPAS operated in full spectral resolution, are referred to as MIPAS(1), while climatologies for 2005–2010, when MIPAS operated in reduced spectral resolution, are referred to as MIPAS(2). SMR provided a second ozone product measured at

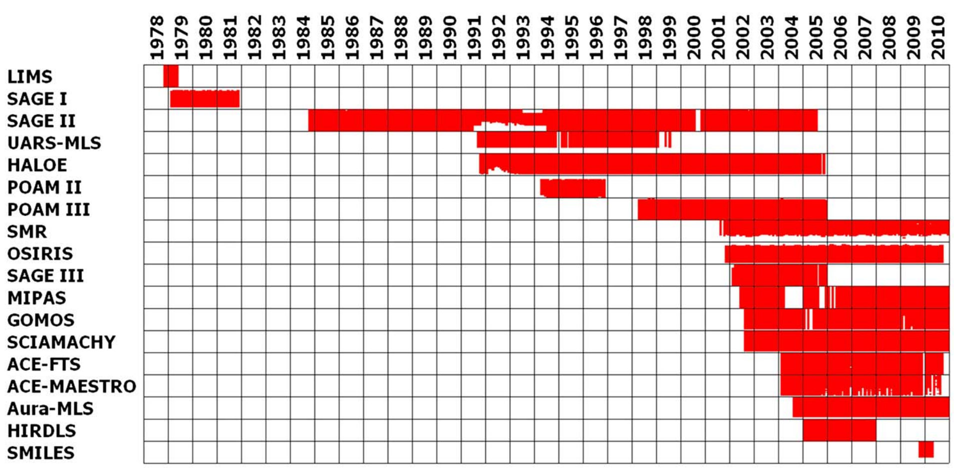

Figure 1. Available ozone measurement records between 1978 and 2010 from limb-sounding satellite instruments participating in the SPARC Data Initiative. The red filling of the grid boxes indicates the temporal (Jan–Dec) and vertical (300 to 0.1 hPa) coverage of the instruments.

Table 2. Data Version, Time Period, Vertical Range and Resolution, and References are Given for Ozone Data Sets Participating in the SPARC Data Initiative

Instrument and Data Version	Time Period	Vertical Range	Vertical Resolution	References
LIMS V6.0	Nov 1978–May 1979	10–80 km	3.7 km	*Remsberg et al.* [2007]
SAGE I V5.9	Feb 1979–Nov 1981	10–55 km	1 km	*McCormick et al.* [1989] *Wang et al.* [1996]
SAGE II V6.2	Oct 1984–Aug 2005	5–70 km	0.5–1 km	*Chu et al.* [1989] *Wang et al.* [2002]
UARS-MLS V5	Oct 1991–Oct 1999	17–75 km	3.5–5 km 5–8 km (>50 km)	*Livesey et al.* [2003]
HALOE V19	Oct 1991–Nov 2005	10–90 km	2.5 km	*Grooß and Russell* [2005]
POAM II V6.0	Oct 1993–Nov 1996	15–50 km	1 km	*Lumpe et al.* [1997] *Rusch et al.* [1997]
POAM III V4.0	Apr 1998–Dec 2005	5–60 km	1.0 km	*Lumpe et al.* [2002] *Randall et al.* [2003]
SMR V2-1	Jul 2001	18–65 km	2.5–3.5 km	*Urban et al.* [2005]
OSIRIS V5-0	Oct 2001	10–60 km	2 km	*Degenstein et al.* [2009]
SAGE III V4.0	Feb 2002–Dec 2005	5–85 km	0.5–1 km	*Wang et al.* [2006]
MIPAS(1) V9	Mar 2002–Mar 2004	6–68 km	3.5–5.0 km	*Steck et al.* [2007]
MIPAS(2) V220	Jan 2005–Apr 2012	6 – 70 km	2.7–3.5 km	von *Clarmann et al.* [2009]
GOMOS V5.0	Aug 2002–Apr 2012	15–100 km	2–3 km	*Kyrölä et al.* [2010]
SCIAMACHY V2.5	Aug 2002–Apr 2012	10–60 km	3–5 km	*Mieruch et al.* [2012]
ACE-FTS V2.2 update	Mar 2004	5–95 km	3–4 km	*Dupuy et al.* [2009]
ACE-MAESTRO V1.2	Mar 2004	5–60 km	2 km	*Dupuy et al.* [2009]
Aura-MLS V2-2	Aug 2004	12–75 km	3 km 4 km (>60 km)	*Froidevaux et al.* [2008] *Jiang et al.* [2007]
HIRDLS V6.0	Feb 2005–Dec 2007	10–55 km	1 km	*Nardi et al.* [2008] *Gille et al.* [2008]
SMILES V2-1-5	Oct 2009–Apr 2010	16–96 km	3–5 km	*Baron et al.* [2011]

488.9 GHz, which has very similar characteristics compared to the main SMR ozone product at 501.8 GHz and is not shown in the following evaluations. For ACE-MAESTRO, the ozone product derived from the visible spectra is used while the UV ozone product is not included. The SMILES products Band-A O_3 and Band-B O_3 show very similar characteristics, and therefore, only one SMILES product (Band-A O_3) is included here.

3.2. Climatology Diagnostics and Uncertainties

[10] This study aims to analyze the mean differences between the various ozone data sets and to identify their vertical, latitudinal, and temporal structure. A set of standard diagnostics including annual and monthly zonal mean climatologies, vertical and meridional mean profiles, seasonal cycles, and interannual variability is used for this purpose. Additionally, the physical consistency of the data sets is tested through diagnostics based on the QBO and the Antarctic ozone hole. Although some instrument retrievals involve constraints that add prior information to the resulting profiles or reduce the altitude resolution, the evaluations presented here are based on direct climatology comparisons without any considerations of averaging kernels. This approach is justified, since for all limb sounders participating in the SPARC Data Initiative, the profiles are well resolved in most parts of the altitude range considered. The notations for different atmospheric regions used throughout the evaluations are given in Table 3.

[11] We will use the multi-instrument mean (MIM) as a common point of reference. The MIM is calculated as the mean of the monthly zonal mean time series from all available instruments within a given time period of interest. We calculate relative differences as the absolute difference of an instrument climatology to the MIM divided by the MIM. It should be stated that the MIM is not a data product and will not be provided as part of the SPARC Data Initiative data set. The choice of the MIM is by no means based on the assumption that it is the best estimate of the atmospheric ozone field but is motivated by the need for a reference that does not favor a certain instrument. Note that the MIM has a number of shortcomings, including the fact that the composition of instruments from which the MIM is calculated can change between different time periods and regions.

[12] Monthly zonal mean ozone climatologies can be affected by the presence of errors in the measurements. While random errors have little impact on climatological means, measurement biases will produce differences between the climatology and the truth. Biases of the raw measurements are related to retrieval errors, uncertainties in the retrieval parameters (e.g., spectroscopic data), and so-called smoothing errors related to spatial resolution of the retrievals. Additionally, monthly mean data contain errors introduced through the climatology production due to instrument sampling [*Toohey et al.*, 2013] and different averaging techniques [*Funke and von Clarmann*, 2012]. The overall errors of the climatologies, which contain the systematic errors of both the measurements and the climatology construction, would allow us to assess the uncertainty in our knowledge of the atmospheric ozone mean state. However, such overall bottom-up errors are not available according to a common standard for all instruments, and therefore, we will use the interinstrument spread of climatologies as a measure of the overall uncertainty in the underlying ozone field.

[13] An approximate measure of uncertainty in each climatological mean is the standard error of the mean (SEM), calculated from n measurements and a standard deviation,

Table 3. Definitions and Abbreviations of Different Atmospheric Regions Used for the Evaluations

Region	Abbreviation	Lower Boundary	Upper Boundary
Upper Troposphere	UT	300 hPa	Tropopause
Lower Stratosphere	LS	Tropopause	30 hPa
Middle Stratosphere	MS	30 hPa	5 hPa
Upper Stratosphere	US	5 hPa	1 hPa

SD, as SEM = $SD/\sqrt{n}$. Note that the SEM could be an overestimate or underestimate of the true uncertainty in the mean since individual samples may exhibit positive or negative autocorrelations [*Toohey and von Clarmann*, 2013]. Despite this shortcoming, due to its frequent use in past studies, the SEM will be used as an approximate measure of uncertainty in each climatological mean, graphically illustrated by 2xSEM error bars, which can be loosely interpreted as a 95% confidence interval of the mean.

3.3. Multiannual Mean Climatology Evaluations

[14] In contrast to the strong, statistically significant ozone decline observed until the mid-1990s (6–8% per decade in the US), ozone trends since then have been near zero or slightly positive [*WMO*, 2011 and references therein]. In order to avoid the impact of the strong trend before the mid-1990s, the multiannual mean climatology evaluations will be based on the time period 1996–2010. Data sets from SAGE II, UARS-MLS, HALOE, POAM II, POAM III, SMR, OSIRIS, SAGE III, MIPAS(1/2), GOMOS, SCIAMACHY, ACE-FTS, ACE-MAESTRO, Aura-MLS, and HIRDLS are included in the 1996–2010 evaluations. Note that none of the instruments covers the full time period. Detailed evaluations of shorter time periods (e.g., 1994–1996, 2003–2004, 2005–2010) give very similar results for the individual instruments available over the respective time period, justifying the approach used here. LIMS and SAGE I data are compared separately for their overlap period in 1979. SMILES data extend from January to April 2010 and are evaluated against the MIM of all instruments available for this time period. Since SAGE II has a very long data record and is used extensively in validation and long-term studies, it is also of interest to use SAGE II as a reference for comparisons with other satellite measurements. These additional comparisons are derived for the maximum overlap time period for each individual instrument with the SAGE II mission.

[15] Evaluations of the multiannual mean climatologies include the comparison of annual mean (pressure-latitude) cross sections and of monthly mean vertical profiles for each instrument. While the cross sections reveal the overall global structure of the mean biases between the different data sets, the profiles are used to focus on particular latitude regions and months. Furthermore, the ozone seasonal cycle is included in the evaluations of the multiannual mean climatologies, revealing to what degree seasonal variations in ozone are captured by the different instruments.

[16] In the mesosphere, day and nighttime differences exist due to photodissociation within the odd oxygen families [e.g., *Brasseur and Solomon*, 1984]. The resulting diurnal ozone variations are of ~10% below 1 hPa and grow with increasing altitude up to more than 100% for upper mesospheric levels [e.g., *Wang et al.*, 1996; *Schneider et al.*, 2005]. Depending on the instruments' sampling pattern, the diurnal cycle may cause systematic biases in the mesosphere and the climatologies are not evaluated above 1 hPa. Note that the impact of temperature uncertainties on the conversion from altitude to pressure during the climatology production may cause additional errors in the US and mesosphere.

3.4. Annual Zonal Mean Cross Sections

[17] Figure 2 shows the annual zonal mean MIM ozone climatology for 1996–2010 and the differences of the individual climatologies with respect to the MIM. The differences are calculated based on multiannual/annual mean data sets and are impacted by the coverage of the satellite instrument in question. If, in a given latitude bin, an instrument does not cover the complete time period, it can be biased toward the months and years when data are available. Due to their limited latitudinal coverage, POAM II/III and SAGE III are not included in the evaluation of the annual mean cross sections.

[18] The tropical and midlatitude MS/US is characterized by the smallest relative differences. The climatologies from SAGE II, UARS-MLS, OSIRIS, GOMOS, Aura-MLS, and HIRDLS yield a very good agreement with differences to the MIM of up to ±5% and often even below ±2.5%. Slightly larger differences are found for HALOE, MIPAS (1/2), and ACE-MAESTRO, reaching values of up to ±5% and in some regions up to ±10%. SMR, SCIAMACHY, and ACE-FTS show a good agreement with the other instruments with positive differences of up to +10% (+20% in the US) for the latter two and negative differences of up to −10% in the case of SMR.

[19] In the LS, differences are larger compared to the regions above; however, in the midlatitudes and tropics, SAGE II, MIPAS (1/2), and Aura-MLS agree well with differences only occasionally exceeding ±10%. Most other instruments agree reasonably well with differences up to ±20%. Exceptions are UARS-MLS, HALOE, OSIRIS, GOMOS, and HIRDLS, which show considerable local disagreement of up to ±50%. While the large differences in the LS are for most instruments only present in the tropics, GOMOS also shows considerable disagreement of up to ±50% at middle and high latitudes. In general, GOMOS differences to the MIM show a strong vertical gradient with good agreement above 70 hPa and a sharp increase in the differences at pressure levels below 70 hPa. UARS-MLS ozone values below 100 hPa have a known high bias and their use is not recommended [*Livesey et al.*, 2003]. While these pressure levels have not been included in the UARS-MLS climatology, the levels directly above 100 hPa can be affected through interpolation of the high-biased values. SAGE II and OSIRIS in the LS show very good agreement with the MIM in Northern Hemisphere (NH) high latitudes but display large deviations at SH high latitudes.

[20] Strong positive deviations of SCIAMACHY from the MIM in the MS/US of up to 20% are possibly related to the fact that before 2006, the SCIAMACHY climatology above 3 hPa suffers from insufficient vertical resolution and coverage of the ECMWF temperature data used to convert originally retrieved number density into VMR. Note that relative differences of SCIAMACHY to the MIM are smaller in 2006–2010 compared to 2003 (not shown here). Deviations of OSIRIS from the MIM vary with latitude, which is most likely caused by sampling biases introduced by nonuniform monthly and yearly sampling. ACE-FTS and ACE-MAESTRO, which also suffer from nonuniform sampling, show differences with respect to the MIM that are very similar in structure but opposite to those of OSIRIS, Aura-MLS, and GOMOS consistent with a validation study by *Dupuy et al.* [2009].

[21] At high latitudes, differences with respect to the MIM are larger (locally up to ±50%) when compared to the tropics and midlatitudes, in particular for OSIRIS in the SH, SAGE

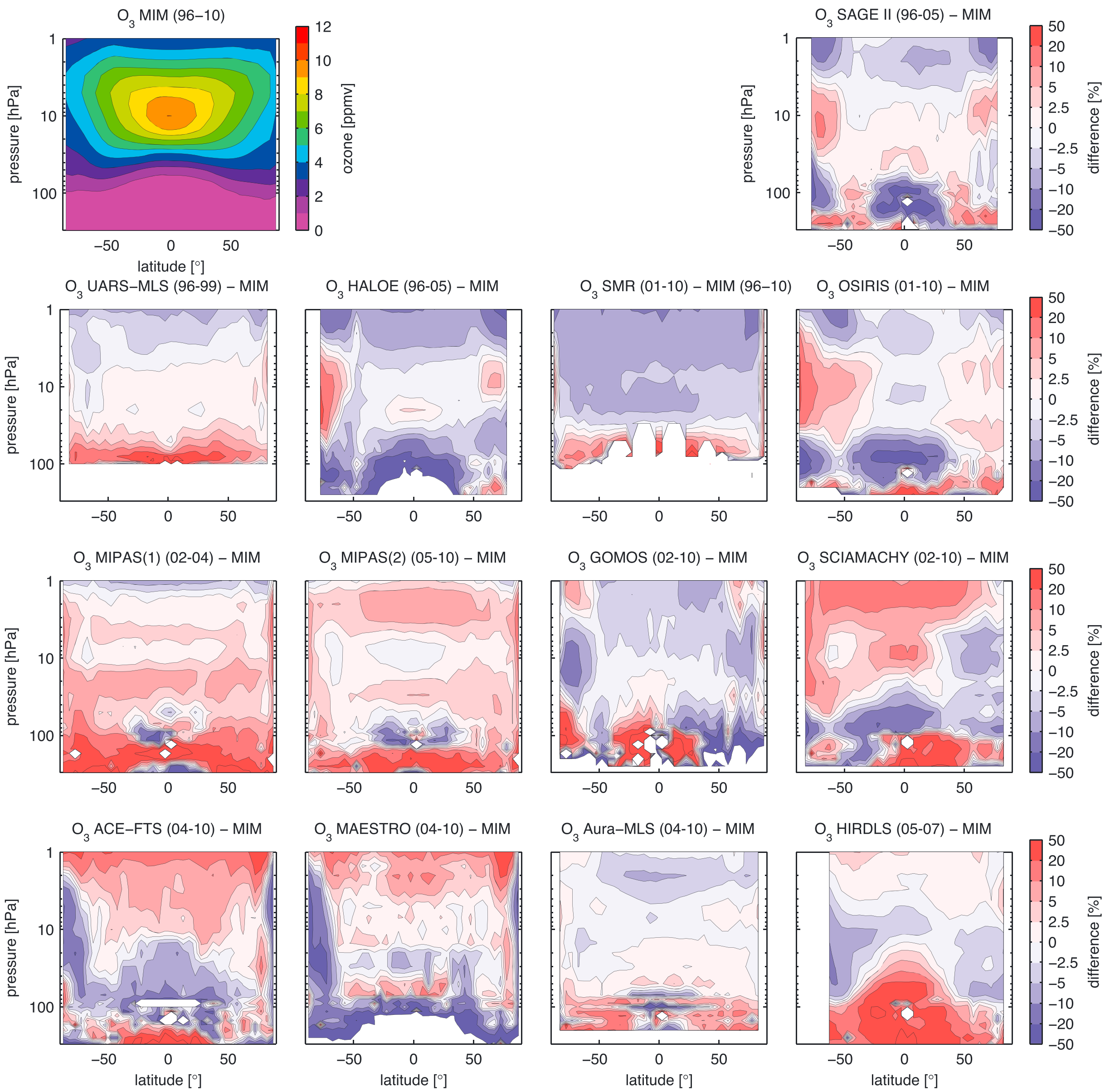

Figure 2. Cross sections of the MIM annual zonal mean ozone for 1996–2010 and differences between the individual instruments and the MIM are shown. The MIM includes SAGE II, UARS-MLS, HALOE, SMR, OSIRIS, MIPAS(1/2), GOMOS, SCIAMACHY, ACE-FTS, ACE-MAESTRO, Aura-MLS, and HIRDLS. Note that while none of the instruments covers the full time period, detailed evaluations of shorter time periods (e.g., 1994–1996, 2005–2010) give very similar results.

II, HALOE, GOMOS, SCIAMACHY, ACE-FTS, and ACE-MAESTRO. The large differences at high latitudes are partially caused by the effects of nonuniform temporal sampling. The annual mean climatologies from instruments with incomplete yearly coverage will be biased toward the months when measurements are available, which produces an especially strong effect in the SH high latitudes up to 20% in some cases [*Toohey et al*., 2013]. However, the large differences at high latitudes observed for some instruments (e.g., SAGE II, HALOE, OSIRIS (SH), and GOMOS in 2003) are also present in the monthly mean comparisons [Figures A1.9 to A1.16 in the SPARC Data Initiative report] and are not exclusively introduced by the annual averaging.

[22] A comparison of the LIMS and SAGE I climatologies for February to April 1979 is shown in Figure 3 (left and middle panels). In the MS, both instruments show excellent agreement, with differences from their MIM mostly within ±2.5% for all latitude bands (corresponding to a direct difference between the two instruments of less than 5%), while differences in the LS are larger, reaching up to ±20%. For all 3 months included in the evaluation, LIMS has mostly negative deviations when compared to SAGE I. Note that the differences are reversed in May when LIMS has a mostly positive deviations from SAGE I [Figures A1.25–A1.26 in the SPARC Data Initiative report], very likely related to SAGE I sampling issues, with sunrise measurements only early in

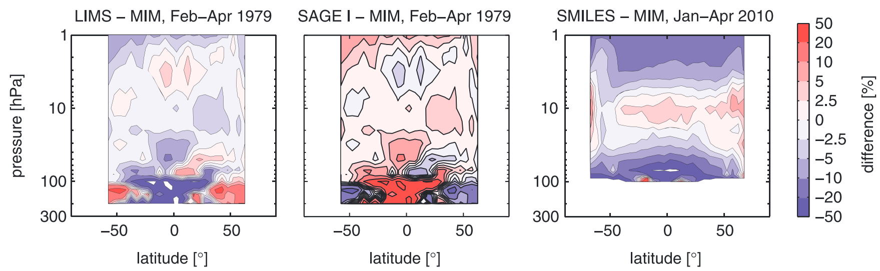

Figure 3. Cross sections of annual zonal mean ozone differences between LIMS and SAGE I and their MIM for (left and middle) Feb–Apr 1979 and between SMILES and the MIM of all climatologies available for (right) Jan–Apr 2010.

the month in the Northern Hemisphere (NH) and sunset measurements only in the SH.

[23] The evaluation of the SMILES climatology is based on a comparison of its zonal mean ozone cross sections (averaged from January to April 2010) to the MIM of all instruments available for the same time period (ACE-FTS, ACE-MAESTRO, Aura-MLS, GOMOS, MIPAS(2), OSIRIS, SCIAMACHY, SMR, and SMILES). The differences are presented in Figure 3 (right panel) and indicate a very good agreement in the MS with deviations of up to +5%. However, in the US, differences of up to −20% are found.

3.5. Monthly Zonal Mean Profiles

[24] The evaluation of the monthly zonal mean profiles is based on typical ozone profiles in the tropics (0°S–5°S in October) and at high latitudes in spring (70°S–75°S in October and 65°N–70°N in April). The location of the latitude bin within each region of interest has been chosen in order to have a maximum number of instruments available. Differences between the MIM profile and the individual instruments are displayed in Figure 4.

[25] In the tropical MS, the monthly mean data sets show the overall best agreement, confirming the results of the annual mean cross sections. All instruments agree within ±5% except for ACE-FTS, which has deviations of up to −10% below 10 hPa, and SCIAMACHY, which has deviations of up to +20%, clearly overestimating the ozone-mixing ratio peak at 10 hPa. In the tropical US, the instruments show good agreement, with the largest deviations of only up to ±15%. The instruments can be divided into two groups with MIPAS(2), SCIAMACHY, ACE-FTS, and ACE-MAESTRO showing a very good agreement among themselves with differences to

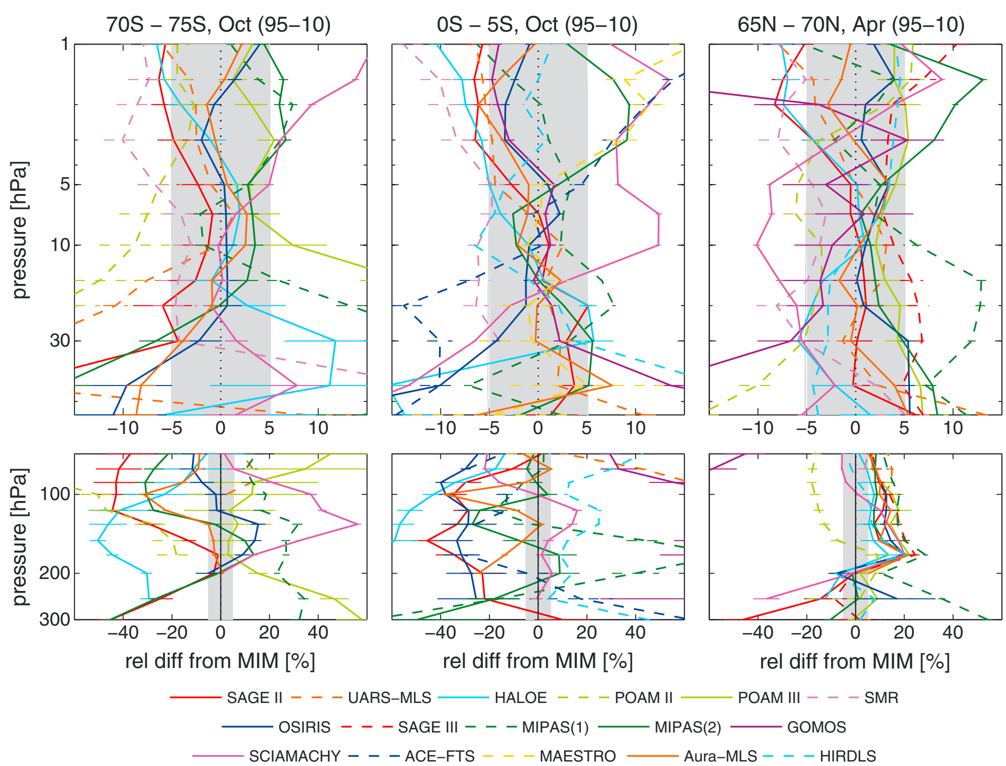

Figure 4. Profiles of monthly zonal mean ozone differences to their MIM for 1996–2010 are presented for 70°S–75°S, 0°S–5°S for October, and 65°N–70°N for April. Bars indicate the uncertainties in each climatological mean based on twice the SEM. The gray-shaded area indicates where relative differences are smaller than ±5%.

the MIM of around +10%, while all other instruments have negative deviations to the MIM of up to −10%. In the midlatitude MS/US, the situation is very similar with the largest spread caused by SMR on the negative side of the MIM, and SCIAMACHY and ACE-FTS on the positive side (not shown here). Overall, the best agreement is found between SAGE II, OSIRIS, GOMOS, and Aura-MLS in the tropical and midlatitude MS/US.

[26] In the tropical LS, differences are large (up to ±50%), as already noted for the annual mean comparisons. This is likely related to the generally lower ozone abundance and the steep vertical ozone gradient in this region that is resolved in different ways by the various instruments. Also, there are instrumental limitations in this altitude region, resulting from, for example, cloud interference and high extinction rates, which can vary depending on the spectral regions and measurement mode and can lead to retrieval errors or to different filtering of the measurements. In particular, HIRDLS, GOMOS, and UARS-MLS show large positive deviations in the LS. Note that the large GOMOS deviations are accompanied by large uncertainties in the climatological mean values.

[27] At high latitudes, all instruments agree very well in the NH and less well in the SH, with differences up to ±20% in the MS. In particular, the very good agreement of the climatologies in the NH high-latitude LS (with differences of ±5% to ±10% except for POAM II and GOMOS) is striking when compared to the SH counterpart where differences are in the range of ±50%. Note that the SEM is also larger at high SH latitudes compared to other regions indicating a higher uncertainty in the climatological mean values. Particularly large differences can be seen for POAM II on the negative side and POAM III on the positive side. The comparison of monthly zonal mean data at high latitudes, where intramonthly, interannual, and zonal natural variability is high, is complicated by the different sampling patterns of the instruments and can cause sampling biases of up to 20% for some cases [*Toohey et al.*, 2013].

3.6. Comparison With SAGE II

[28] SAGE II has a long data record and shows in general very good agreement when compared to ozonesondes [*Wang et al.*, 2002]. In order to extend trend analysis of global ozone profiles beyond the end of the SAGE II data record in August 2005, information from more recent satellite instruments is needed. The compilation of a homogeneous ozone profile record needs to account for the small shifts between the various satellite time series that inevitably occur. In order to determine which instruments show the best agreement with SAGE II in each region and could therefore be a suitable choice for data merging activities, a comparison to SAGE II is performed here. The comparisons are done for the maximum overlap time period of each individual instrument with SAGE II, i.e., each comparison is based on a different time period varying from 15 years (for the comparison with HALOE) to 6 months (for the comparison with HIRDLS).

[29] Figure 5 displays the comparison of SAGE II to 15 instrumental climatologies. Deviations between SAGE II and the individual instruments vary with latitude and altitude, and no instrument can be singled out as giving the best agreement with SAGE II everywhere. In the tropical and midlatitude MS, GOMOS and Aura-MLS show excellent agreement with differences below ±2.5%, while UARS-MLS, HALOE, OSIRIS, SAGE III, and MIPAS(1) have only slightly larger deviations to SAGE II, often up to ±5%. The largest departure from SAGE II can be found for ACE-MAESTRO, ACE-FTS, SCIAMACHY, and MIPAS(2) with differences up to ±20% in some places. The latter has already been identified in earlier MIPAS versions [*Stiller et al.*, 2012] and thus indicates a problem in the MIPAS level 1 data. For the tropics and midlatitude US, the comparisons show similar results as in the MS, with excellent agreement of OSIRIS and GOMOS to SAGE II (±2.5%).

[30] In the tropical LS, differences to SAGE II data increase with decreasing altitude for most instruments. Aura-MLS, OSIRIS, as well as MIPAS(1/2) display the best agreement. In the tropical UT, nearly all data sets (except HALOE and ACE-MAESTRO) show larger ozone values than SAGE II, consistent with the known low bias for SAGE II with respect to ozonesondes in this region [*Wang et al.*, 2002]. Observed vertical oscillations (mainly at low latitudes) in the Aura-MLS panel are largely caused by systematic oscillatory features in the MLS UTLS retrievals that are expected to be mitigated in a future data version. In the NH polar latitudes, HIRDLS, UARS-MLS, OSIRIS, and MIPAS(1) agree well and POAM III agrees very well with SAGE II. In the SH, Aura-MLS, OSIRIS, HALOE, SMR, and UARS-MLS have only small offsets compared to SAGE II of up to ±10%, while other instruments reveal larger differences of up to ±20% or even ±50% in the case of GOMOS.

[31] Overall, more than half of the instruments agree very well with SAGE II in the MS/US showing mean deviations of less than 5%. Note that some differences between the climatologies can be accounted for by differences between the instrumental absorption cross sections. For example, the ozone cross section used in the SAGE II retrieval is about 2% lower compared to the one used by GOMOS. Neglecting other potential systematic differences, we would then expect SAGE II to be about 2% larger than GOMOS due to the different ozone cross sections, which is in fact the case in the MS. Above and below the MS/US, a large spread of the climatological deviations can be found with differences as small as 10–20% or as large as 50%. The overall best agreement to SAGE II is found for Aura-MLS, OSIRIS, and MIPAS(2) in the LS, Aura-MLS and GOMOS in the MS, and OSIRIS and POAM II in the US.

4. Seasonal Cycle

[32] The evaluation of the seasonal cycles is based on the multiannual mean approach and is used to determine if biases between data sets are persistent over the entire year. The seasonal cycle plots (Figure 6) include the interinstrument standard deviation, which acts as a measure of the range of mean values given by the different instruments. For each instrument, a combined annual and semiannual fit has been applied to all the available monthly mean values. The derived fit for the seasonal cycle is shown by the lines while the individual data points are represented by the symbols.

[33] Ozone above 10 hPa exhibits a strong semiannual cycle associated with the tropical semiannual oscillation in zonal wind and temperature [*Ray et al.*, 1994]. Figure 6 (upper left panel) shows how well the seasonal cycle of tropical

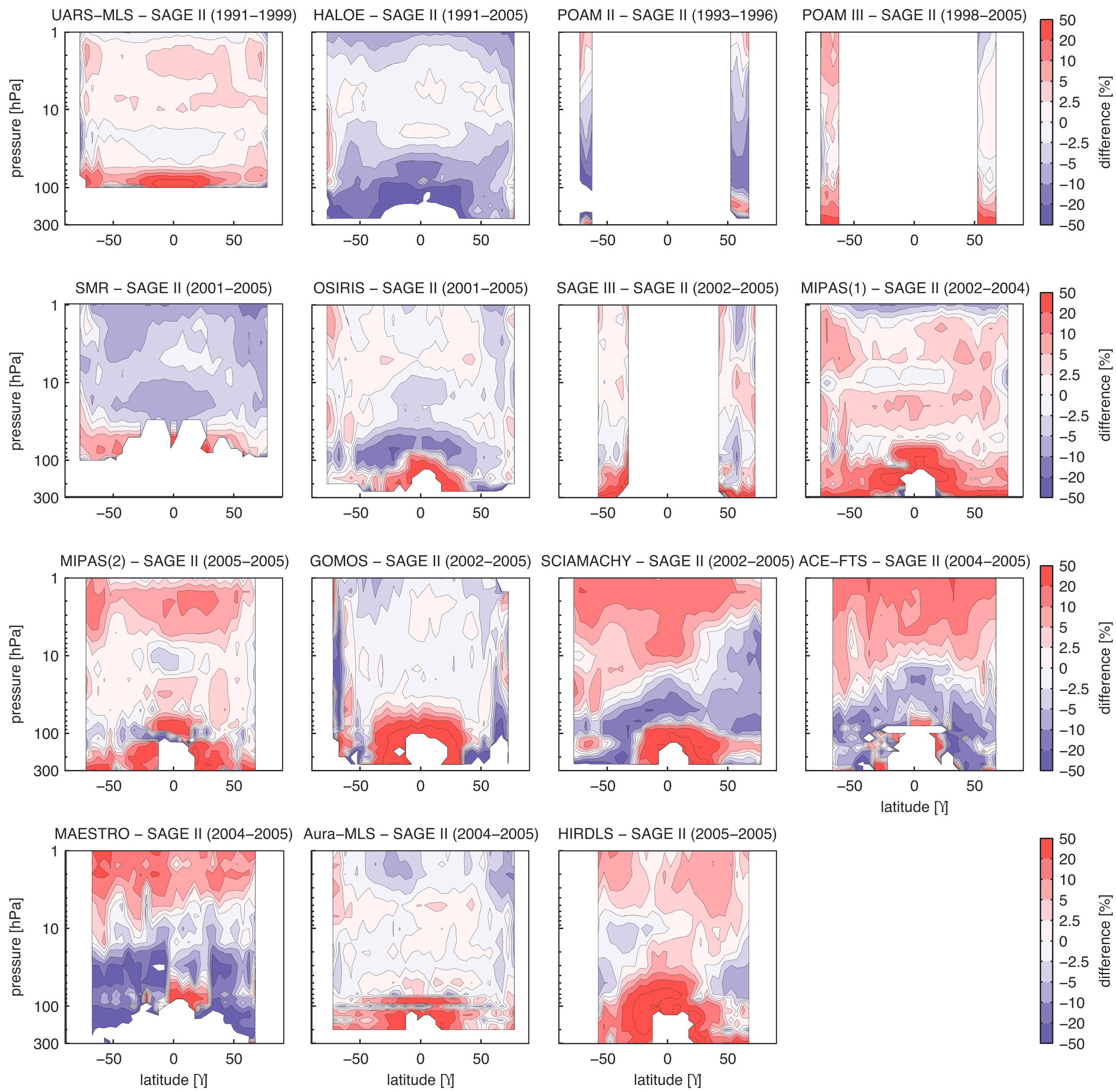

Figure 5. Cross sections of zonal mean ozone differences to SAGE II. Zonal mean ozone differences between the individual instruments and SAGE II are shown for time periods of maximum overlap.

(20°S–20°N) ozone at 10 hPa is captured by the individual instruments. All climatologies display the general structure of the semiannual cycle, which is characterized by stronger amplitude during the first half of the calendar year. The seasonal cycle is well captured by SAGE II, HALOE, GOMOS, MIPAS(2), and Aura-MLS, which show only small differences in phase and amplitude when compared to the MIM. SMR, SCIAMACHY, and OSIRIS display the same phase as this group of instruments but have smaller (SMR) or larger (SCIAMACHY and OSIRIS) amplitudes. While HIRDLS agrees quite well during the first half of the calendar year, its seasonal cycle in the second half of the year is too low in amplitude and the mean values are too small. Due to their limited temporal sampling in the tropics, ACE-MAESTRO and ACE-FTS climatologies provide only weak constraints for fitting a seasonal cycle. In particular, for ACE-MAESTRO, a much higher than expected June value prevents fitting a seasonal cycle, although the other monthly mean values are very close to the MIM whenever they are available. UARS-MLS agrees quite well for most months but shows an outlier for May, which adversely affects the fit.

[34] A large annual cycle in tropical ozone near and above the tropopause has been identified from ozonesonde measurement records [e.g., *Randel et al.*, 2007]. The signal extends over only a narrow vertical range, from approximately 100 to 50 hPa, and is related to seasonal changes in vertical transport acting on the strong vertical ozone gradient in this region. Since it can be used to analyze the seasonal changes in tropical upwelling, the seasonal cycle is an important characteristic of tropical ozone in the LS. Figure 6 (lower left panel) demonstrates that the satellite instruments have difficulties in estimating the ozone seasonal cycle at 80 hPa and large differences in the projected amplitude and phase can be observed. UARS-MLS shows significantly larger ozone values compared to the other instruments [*Livesey et al.*, 2003], as already noted for the annual mean comparison. Despite this offset, UARS-MLS, SAGE II, HALOE, and Aura-MLS estimate a very similar amplitude and phase for the seasonal cycle with

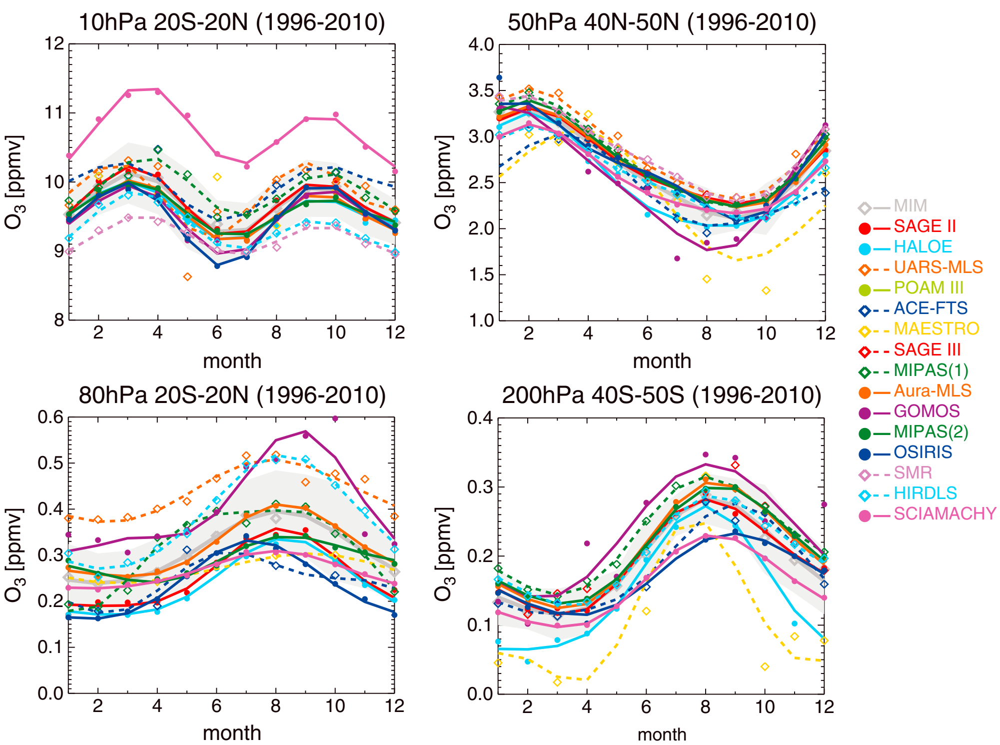

Figure 6. Seasonal cycle of monthly zonal mean ozone for 20°S–20°N at 10 hPa and 80 hPa, for 40°N–50°N at 50 hPa, and for 40°S–50°S at 200 hPa for 1996–2010. The gray shading indicates the MIM ± 1σ multi-instrument standard deviation.

maximum values in July or August. All other instruments also show elevated values in NH summer; however, there is no agreement between the instruments regarding the amplitude or phase of the annual cycle. The instrument-by-instrument analysis reveals that while SCIAMACHY and HIRDLS have a very similar phase compared to the MIM, they show a much smaller or, respectively, a much larger amplitude, with the latter possibly related to HIRDLS' better vertical resolution. GOMOS, MIPAS(1), OSIRIS, MIPAS(2), ACE-FTS, and ACE-MAESTRO all show considerable differences in the phase, and additionally, the first two display a too large and the latter three a too small seasonal cycle amplitude. These interinstrument inconsistencies are related to the strong vertical gradient in ozone in this range and the narrow vertical region over which the annual cycle extends [*Randel et al.*, 2007]. As a result, mean values and seasonal variability are quite sensitive to the vertical resolution and sampling characteristics of the measurements.

[35] The ozone seasonal cycle in the NH midlatitude MS, with a maximum in early spring and a minimum in late summer/fall, is related to transport variations of the large-scale stratospheric circulation. At 50 hPa, the absolute mean values and the annual cycle agree very well between all instruments (upper right panel in Figure 6). Exceptions are GOMOS and ACE-MAESTRO, which both show values well beyond the 1σ range for some parts of the year as well as different phases and larger amplitudes of the seasonal cycle compared to other instruments. The ozone seasonal cycle in the UTLS in SH midlatitudes has a maximum in SH late summer/fall, resulting from large-scale transport processes and their seasonal variations (lower right panel in Figure 6). There is a large spread in the signal displayed by the instruments, with the biggest discrepancies observed for OSIRIS, HALOE, and ACE-MAESTRO with larger amplitudes for the latter two and a relatively shallow seasonal cycle for OSIRIS. The evaluation of the NH seasonal cycle at 200 hPa (not shown here) gives a better agreement, with nearly all instruments including OSIRIS showing a consistent seasonal cycle.

5. Interannual Anomalies

[36] Time series of deseasonalized anomalies are used to analyze interannual ozone variability, which is related to a number of chemical and dynamical processes. These processes include the QBO signal, variations of the Brewer Dobson circulation, the solar cycle, volcanic perturbations of stratospheric aerosols, and the variability of the polar vortex strength. The evaluation of interannual anomalies helps to understand how well the sensitivity of ozone abundance to various processes is captured by the individual instruments. The multiannual mean values are based on all years of the evaluation period (2000–2010) available for the respective instrument. For each instrument and month, the anomalies are calculated by subtracting the multiannual monthly mean value from the respective monthly mean values. No additional adjustments are applied to correct the effect of different lengths of the underlying time series.

[37] The QBO is the dominant source of interannual variability in equatorial ozone, and a realistic characterization of the altitude-time QBO structure by satellite measurements is an important aspect of the physical consistency of the data set. The QBO signal in ozone exhibits a double-peaked structure in the vertical, with one maximum in the LS resulting

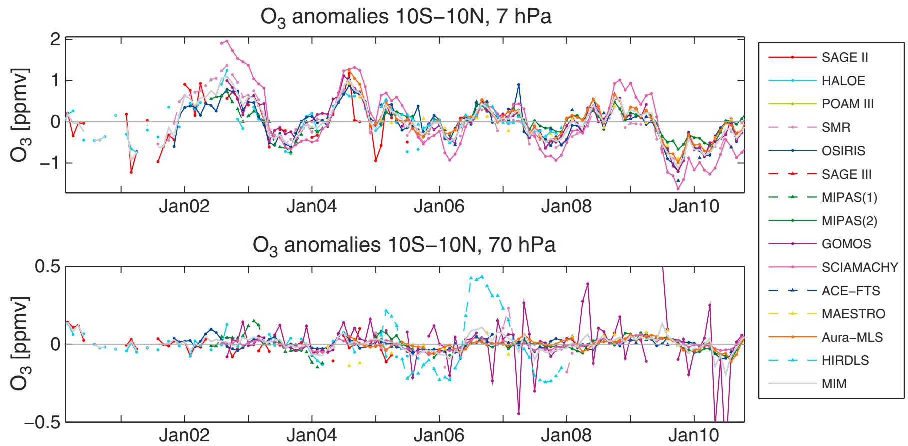

Figure 7. Time series of deseasonalized ozone anomalies at 7 and 70 hPa between 10°S and 10°N for 2000–2010.

from the transport of ozone from the QBO-induced residual circulation [*Zawodny and McCormick*, 1991], and the other maximum in the MS/US arising from QBO-induced temperature variations [*Ling and London*, 1986] together with QBO-induced NO_y variability [*Chipperfield et al.*, 1994]. Figure 7 displays tropical (10°S–10°N) interannual ozone anomalies at 7 and 70 hPa. The QBO signal has an approximate 2 year long cycle, which is well captured at 7 hPa by all instrumental climatologies, although some differences in the amplitude exist. Aura-MLS is in the middle of the range given by all instrument climatologies. Deviations of GOMOS or OSIRIS from the other instruments last only a few months and are independent of the QBO phase. In contrast, MIPAS(2) and SCIAMACHY deviations from the MIM last over longer time periods (up to 2 years) and are related to the QBO phase. While MIPAS(2) ozone anomalies have a lower amplitude, SCIAMACHY shows the opposite behavior with larger positive ozone anomalies and smaller negative anomalies than the other climatologies.

[38] QBO ozone anomalies propagate downward in time, and evaluations of the various pressure levels below 7 hPa (not shown here) give very similar results. At 10 hPa and below, SAGE II displays stronger month-to-month fluctuations than the other instruments. In general, most of the instruments agree better below 15 hPa where ozone is under dynamical control. The deviations of MIPAS(2) and SCIAMACHY to the MIM propagate downward in phase with the underlying QBO ozone signal. The amplitude of the tropical ozone QBO has a maximum around 30 hPa and decreases with decreasing altitude so that at the 70 hPa level, only small amplitudes are observed by the instruments. The only exception to this is HIRDLS, which displays a considerably larger amplitude of the QBO oscillation for the 3 years of available data. Note that ozonesonde data in the tropical LS also indicate a very weak signal of the ozone QBO variations [*Witte et al.*, 2008]. GOMOS reveals large spikes in the time series that are also found for other latitude bands and pressure levels below 30 hPa. Interannual ozone anomalies

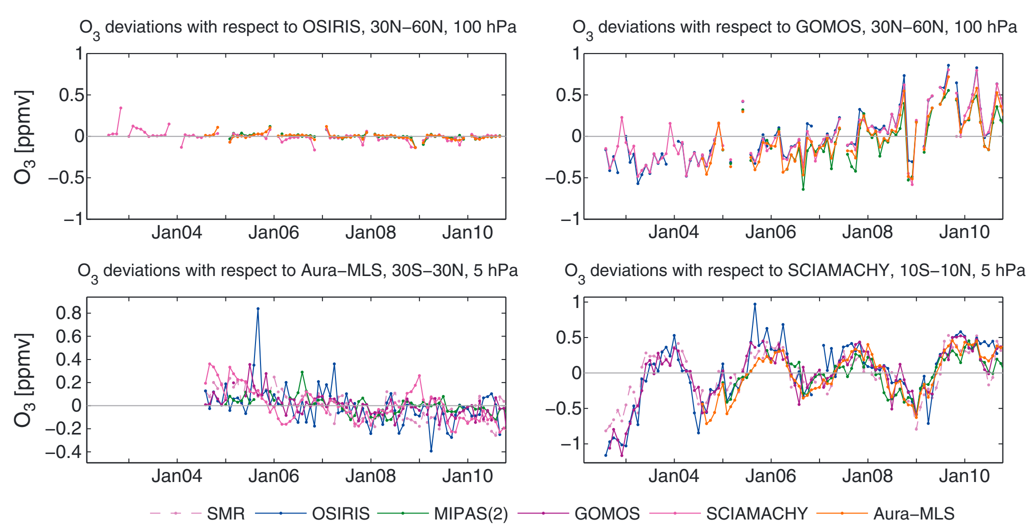

Figure 8. Time series of ozone deviations for 2002–2010. Deviations are shown with respect to (top) OSIRIS and GOMOS for 30°N–60°N at 100 hPa, with respect to (bottom, left) Aura-MLS for 30°S–30°N at 5 hPa and with respect to (bottom, right) SCIAMACHY for 10°S–10°N at 5 hPa.

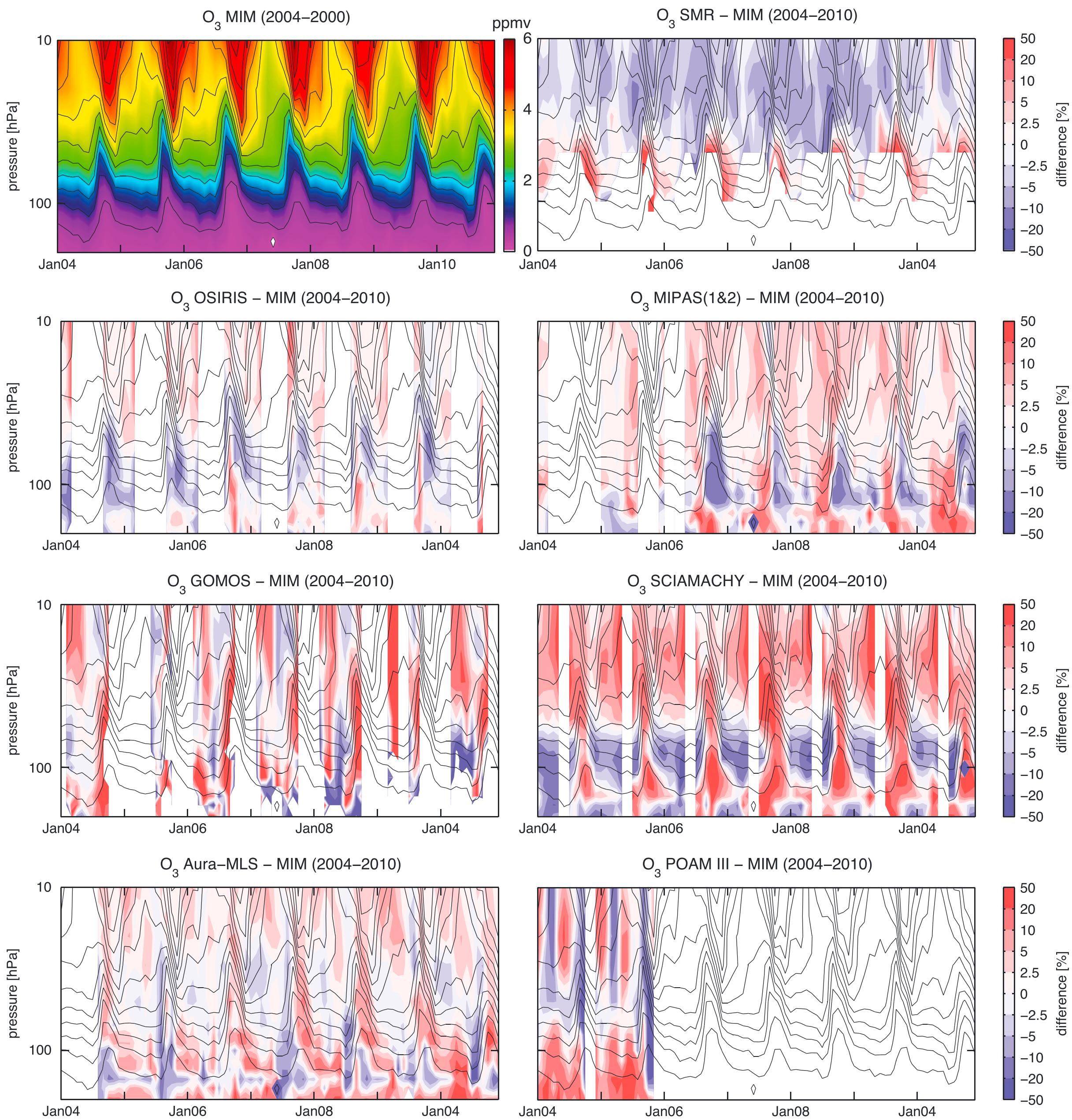

Figure 9. Altitude-time cross sections of MIM ozone (calculated based on displayed data sets) for 60°S–90° S from 2004 to 2010 are shown in upper left panel. Differences between the individual data sets and the MIM are shown in the other panels by color contours. The black contours repeat the MIM ozone field from the upper left panel. Note that the MIM contours are shown for entire time series irrespective of the individual instrument's coverage.

in the extratropical MS (not shown here) give the best agreement for Aura-MLS, HALOE, MIPAS(2), OSIRIS, and SCIAMACHY in the NH subtropics and for Aura-MLS, SAGE II/III, and SMR at NH high latitudes. Here very large anomalies can be observed for individual months for ACE-MAESTRO, GOMOS, or SCIAMACHY, which are not reflected by the other data sets.

[39] Mean differences between the data sets can change over time. For each instrument, an analysis of the time dependence of the differences to each of the other instruments has been performed. Such time series are characterized by seasonal patterns and month-to-month variability. After removing the seasonal cycle, longer-term changes can be the dominant signal. However, for nearly all data sets and regions included in this study, the differences display no apparent long-term changes. One example for this consistency is shown in Figure 8 (upper left panel) in the form of the instrument differences with respect to OSIRIS in the NH midlatitude LS. A few exceptions exist where clear changes of the differences over time can be identified (Figure 8). First, differences of all instruments with respect to GOMOS in the NH midlatitude LS are mostly negative before 2008 and

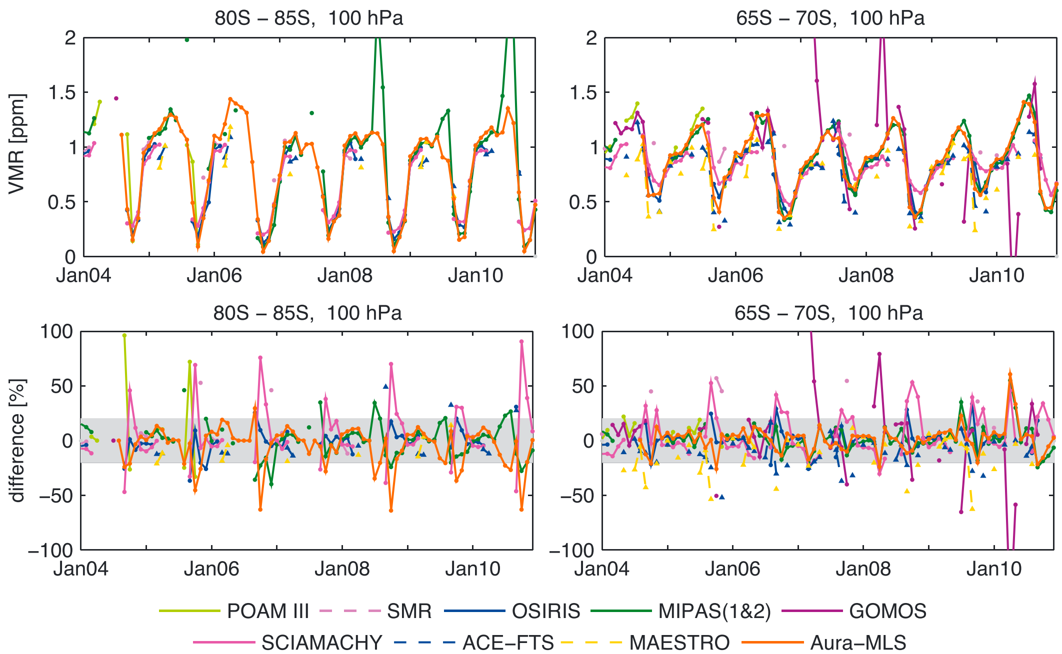

Figure 10. Time series of (top) zonal monthly mean ozone and (bottom) relative differences with respect to MIM at (left) 80°S–85°S and (right) 65°S–70°S for 100 hPa are shown. The gray-shaded area indicates where relative differences are smaller than ±20%.

mostly positive afterwards, indicating a change of GOMOS data over time that is not seen by the other instruments. Note that GOMOS is excluded from the comparison to OSIRIS discussed above in order to present one example where the differences display no apparent long-term changes. For Aura-MLS, a similar change can be observed for the tropical US, with positive differences at the beginning and negative differences at the end of the time period. SCIAMACHY differences in the tropics are dominated by the QBO signal, while SMR (not shown here) displays larger values compared to the other data sets in 2003 but differences around zero after 2006. Note that here, only drifts of a magnitude comparable to the deviations themselves have been identified, while for trend studies, a more thorough analysis including possibly quite small long-term drifts is necessary.

6. Antarctic Ozone Hole

[40] Stratospheric ozone depletion at polar latitudes through catalytic chemistry has been one of the major environmental issues of the last decades [e.g., *Solomon*, 1999; *WMO*, 2011]. Ozone depletion in the polar LS is linked to the activation of chlorine from long-lived reservoir species into reactive forms on the surfaces of polar stratospheric clouds [*Solomon et al.*, 1986; *Molina and Molina*, 1987]. Figure 9 (upper panel) shows the MIM altitude-time cross section of monthly zonal mean ozone averaged over 60°S–90°S (referred to as the polar cap average in the following) from 2004 to 2010. The MIM demonstrates the near complete removal of ozone in the lower stratosphere during Antarctic late winter/early spring as observed by the satellite instruments. In the Antarctic, reactive chlorine can be present for 4–5 months [e.g., *Santee et al.*, 2003], leading to severe ozone depletion in the lower stratosphere as displayed in Figure 9 and thereby reducing total column ozone by as much as two thirds [*WMO*, 2011]. At the end of the year, the ozone hole disappears as a result of the increasing polar stratospheric temperatures and the exchange of air between polar and midlatitudes. Also visible in the ozone altitude-time section is the diabatic descent of air with higher ozone-mixing ratios from the US during winter and spring.

[41] The relative differences between the MIM and the individual instruments for the time evolution of polar cap Antarctic ozone are displayed in Figure 9. The instruments show a considerable disagreement, which is especially pronounced during the time of the Antarctic ozone hole when the mixing ratios are low (as indicated by the underlying MIM ozone field) and when temporal and spatial gradients are strongest. Figure 10 shows the O_3 time series for the individual instruments and their relative differences to the MIM at 100 hPa for the two latitude bins 80°S–85°S and 65°S–70°S. The breakdown of the polar cap average into individual latitude bins allows for the quantification of how much the large differences mentioned above are caused by spatial sampling effects (i.e., for some instruments the polar cap average does not include all latitude bins).

[42] A reasonably good agreement is found between Aura-MLS, MIPAS(1/2), and OSIRIS with polar cap average differences from the MIM of up to ±20%. Aura-MLS (OSIRIS) observes mostly higher (lower) ozone values except during very short periods around the onset of the ozone hole. MIPAS(1/2) differences to the MIM are negative during the time of the ozone hole and positive during the rest of the year. These characteristics are generally confirmed by the comparisons performed for the individual latitude bins with some exceptions found for particular cases. In the higher latitude bin (80°S–85°S) at 100 hPa, in a few cases, Aura-MLS shows larger deviations to the MIM in the range −50%, while differences for the level above and below (not shown here) are in the range ±20%. Larger deviations (up to ±50%) can also be

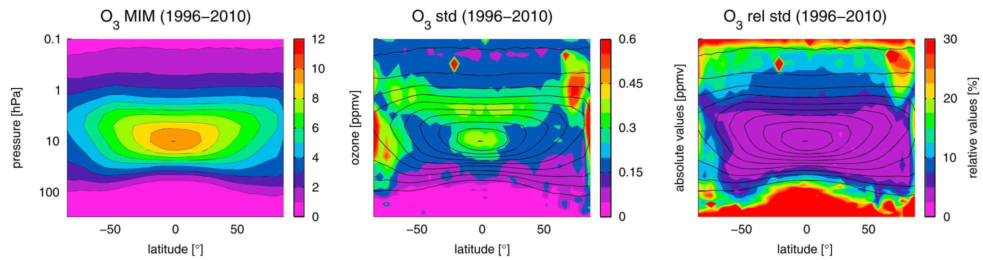

Figure 11. A summary of ozone annual zonal mean state for 1996–2010 is provided in form of the (left) MIM and (middle) absolute and (right) relative standard deviations over all instruments. Relative standard deviations are calculated by dividing the absolute standard deviations by the MIM. Black contour lines on the middle and rightmost panels reproduce the MIM distribution shown in the leftmost panel. Instruments included are SAGE II, UARS-MLS, HALOE, SMR, OSIRIS, MIPAS, GOMOS, SCIAMACHY, ACE-FTS, ACE-MAESTRO, Aura-MLS, and HIRDLS.

found for OSIRIS between 30 and 80 hPa at 80°S–85°S. Note that MIPAS shows exceptionally large inner vortex zonal means at 100 hPa just before the start of the O_3 hole which are visible as peaks in the VMR time series for the years 2007 to 2010 and lead to differences compared to Aura-MLS of up to 50%.

[43] GOMOS, POAM III, SCIAMACHY, and SMR polar cap averages show considerable disagreement with differences up to ±50% and sometimes exceeding ±100%. POAM III and SCIAMACHY polar cap differences to the MIM are linked to the seasonal cycle, with enhanced differences in winter and spring. POAM III observes more ozone than most other instruments (+20%) except during the peak of the ozone depletion at the end of winter when it underestimates the ozone abundance (−50%), while SCIAMACHY deviations are of opposite sign. The analysis of the two latitude bins reveals that POAM III agrees reasonably well with the MIM in the outer vortex (±20%) but shows large deviations in the inner vortex, which can be either positive or negative depending on the month and latitude bin. For SCIAMACHY, the deviations in the outer vortex area are mostly below ±50% but can be as large as ±100% in the inner vortex. GOMOS deviations to the MIM are not coupled with the seasonal cycle and the appearance of the ozone hole. The polar cap average picture shows large deviations for GOMOS in all months. For the upper levels (above 80 hPa), this seems to be an artifact resulting from the averaging process since the evaluation of the individual latitude bins reveals small deviations (mostly within ±20%). However, for levels below 80 hPa, deviations become very large, exceeding ±100%. SMR shows small deviations to the MIM during times with no ozone depletion (smaller than ±20%) and large positive deviations during the Antarctic ozone hole (up to +100%). ACE-FTS and ACE-MAESTRO have very limited sampling over the polar cap, and therefore, the comparison of individual latitude bins is more representative than their polar cap average (not shown here). For both instruments, relative differences are enhanced during times of ozone depletion with large positive deviations found for the inner vortex latitude bins (80°S–85° S) and large negative deviations in the outer vortex latitude bins (65°S–70°S).

[44] For most of the instruments, the deviations from the MIM change sign during the time of the ozone depletion and are opposite to their signal during the rest of the year. The polar cap average ozone deviations are influenced by the sampling patterns of the individual instruments and are in some cases (e.g., GOMOS at levels above 80 hPa) larger than differences derived for individual latitude bands. Overall, however, deviations similar to the ones found for the polar cap average ozone field are apparent in 5° wide latitude bins that are completely inside the polar vortex over several months and therefore should be less affected by spatial sampling effects. Such sampling effects can result from nonuniformity in day-of-month sampling with differences in the individual latitude bins of up to ±10% and in some instances up to 30% [*Toohey et al.*, 2013]. The estimates of the sampling bias in the 80°S–85°S are qualitatively similar to the results shown here for a number of instruments, e.g., the positive bias for POAM III in the SH spring. Note that the magnitude of the large relative differences observed during the time of severe ozone depletion is partially related to the low ozone abundance. But in addition to the effect of the low background ozone on the relative differences, the absolute differences themselves are enhanced during the time of the ozone hole as demonstrated by the evaluation of the ozone time series.

7. Summary and Conclusions

[45] A comprehensive comparison of ozone profile climatologies from 18 satellite instruments has been carried out. Overall findings on the systematic uncertainty in our knowledge of the ozone mean state and important characteristics of the individual data sets are presented in the following summary including two synopsis plots.

7.1. Atmospheric Mean State

[46] An estimate of the uncertainty in our knowledge of the ozone atmospheric mean state is derived from the spread between the data sets and presented in the first summary plot (Figure 11). Annual zonal MIM is presented for the main evaluation period for 1996–2010. The spread between the instrumental climatologies is given by the standard deviation over all instruments presented in absolute and relative values to provide a measure of the overall uncertainty in the underlying ozone field. The evaluation of monthly zonal mean ozone climatologies from various limb-viewing satellite instruments shows that the uncertainty in our knowledge of the atmospheric ozone annual mean state is smallest in the tropical MS and midlatitude LS/MS. The evaluation reveals

a 1σ multi-instrument spread in this region of less than ±5%. Maximum ozone-mixing ratios are found in the tropical MS around 10 hPa. Here the absolute values of the various climatologies show the largest spread for the tropical and extratropical stratosphere, with values varying between 10 and 12 ppmv. In the tropical LS, the spread between the data sets increases quickly with decreasing altitude reaching ±30% at the tropical tropopause. In the midlatitude LS, where the average ozone values are similar to those at the tropical tropopause, the various data sets show closer agreement regarding the ozone mean state, with a 1σ of ±10%. At polar latitudes, the climatologies give a larger spread of the ozone mean state (1σ of ±15%) compared to lower latitudes (1σ of ±5%). Maximum variations (up to 1σ of ±30%) are found in the Antarctic LS, resulting from large relative differences in the observations of the ozone hole.

7.2. Performance by Region

[47] Specific interinstrument differences measured as monthly mean deviations of the instrument climatologies with respect to the MIM are estimated for different regions and presented in the second summary plot (Figure 12). For each instrument and region, the deviation to the MIM is given by the median (mean) difference over all grid points in the region and time period. Additionally, for each instrument, the spread of the differences over all grid points in this region is presented. Note that both pieces of information (mean deviation and regional spread) are important for a meaningful assessment of interinstrument differences. The spread over all grid points in a selected region (sample x) is calculated as the standard deviation and median absolute deviation (MAD) which is defined as

$$MAD = median(|x - median(x)|) \tag{1}$$

[48] The MAD represents the interval around the median that contains 50% of the data [*Rousseeuw and Croux*, 1993]. The selected regions consist of the tropics (20°S–20°N) and midlatitudes (30°S–60°S and 30°N–60°N) and four different altitude regions from the UT up to the US between 300 and 1 hPa following the classification given in Table 3.

[49] The middle stratosphere (30–5 hPa) is characterized by the lowest spread between the instrument data sets. In the tropical and midlatitude MS, nearly all instruments show very good agreement with relative differences smaller than ±5%. Exceptions are SMR with negative deviations to the MIM of around $-5 \pm 2\%$ (regional mean ± 1 sigma) and SCIAMACHY in the tropics with positive deviations of around $+5 \pm 5\%$. Note that some data sets (e.g., SCIAMACHY, ACE-FTS, SMILES) show relatively large standard deviations and MADs indicating a wider regional spread of the relative differences while other instruments (e.g., SMR, Aura-MLS) have small standard deviations indicating a narrow distribution of the relative differences around their mean. Such narrow distributions together with small mean difference indicate the excellent agreement (differences < ±2.5%) between these data sets (e.g., OSIRIS, GOMOS, Aura-MLS). In the polar regions, all instruments display larger relative differences compared to lower latitudes, with differences of up to ±20% in the Antarctic and up to ±10% in the Arctic.

[50] In the lower stratosphere (100–30 hPa), there is a clear difference between the performance of the instruments in the tropics and midlatitudes, with much better agreement in the midlatitudes. Here average differences are mostly in the range of ±10%, except for SMILES with an average deviation of −15%. For some instruments, a relatively wide regional spread (over all LS midlatitude grid points) of the differences is found, indicating individual monthly mean differences larger than +20% for UARS-MLS, SMR, and GOMOS and lower than −30% for GOMOS and SMILES. In the tropics, the interinstrument differences are considerably larger and instruments agree only reasonably well, with average differences below ±20% except for UARS-MLS (+25%), HIRDLS (+35%), and SMILES (−30%). For some instruments in the tropics, a large regional spread is found reaching values below −40% for GOMOS and SMILES and well above +40% for UARS-MLS, GOMOS, and HIRDLS. The poor agreement of the mean values and the larger spread are related to the small ozone abundances in this altitude region and instrumental limitations (e.g., resulting from cloud interference). Note that SMR, MIPAS, and Aura-MLS show excellent agreement with differences to the MIM of less than ±5%. Very close to each other with interinstrument differences of less than 5% are SAGE II, SCIAMACHY, ACE-FTS, and ACE-MAESTRO (mean deviations of ~ −10%) as well as HALOE and OSIRIS (mean deviations of ~ −20%). At high latitudes, differences are mostly in the range of ±30% for the SH and ±10% for the NH, similar to the MS.

[51] In the upper troposphere/lower stratosphere (300–100 hPa), most instruments achieve good agreement in the midlatitudes (average differences up to ±10%) with two small exceptions (±15% for HALOE and MIPAS(1)) and two outliers (−40% for ACE-MAESTRO and −80% for SMILES). Large regional spreads of up to ±75% exist for GOMOS, ACE-MAESTRO, and SAGE III. The good agreement observed at midlatitudes is not found in the tropics, where most instruments show differences of ±20% or larger. Maximum deviations are observed for HALOE, UARS-MLS, GOMOS, and ACE-MAESTRO (with average differences beyond ±60%). Nearly all data sets have a large regional spread with maximum values above ±100%.

[52] In the upper stratosphere (5–1 hPa), similar differences between the data sets exist in the tropics and at midlatitudes. In both regions, SAGE II, UARS-MLS, OSIRIS, SAGE III, MIPAS(1), GOMOS, Aura-MLS, and HIRDLS data sets agree very well with average difference around ±5%. Data sets on the low side, with average deviations around −10%, are HALOE, SMR, and SMILES, while data sets on the high side with average deviations around +10% are MIPAS(2), SCIAMACHY, ACE-FTS, and ACE-MAESTRO.

7.3. Instrument Specific Conclusions

[53] The comparison of ozone profile climatologies from 18 different instruments reveals good agreement for most of the data sets in aspects such as mean biases and interannual variability. Depending on latitude and altitude, individual data sets have been identified as outliers or as showing unphysical behavior. In general, no data set is found to be problem free and identified strengths and weaknesses are listed below. Other limitations and caveats, which are not

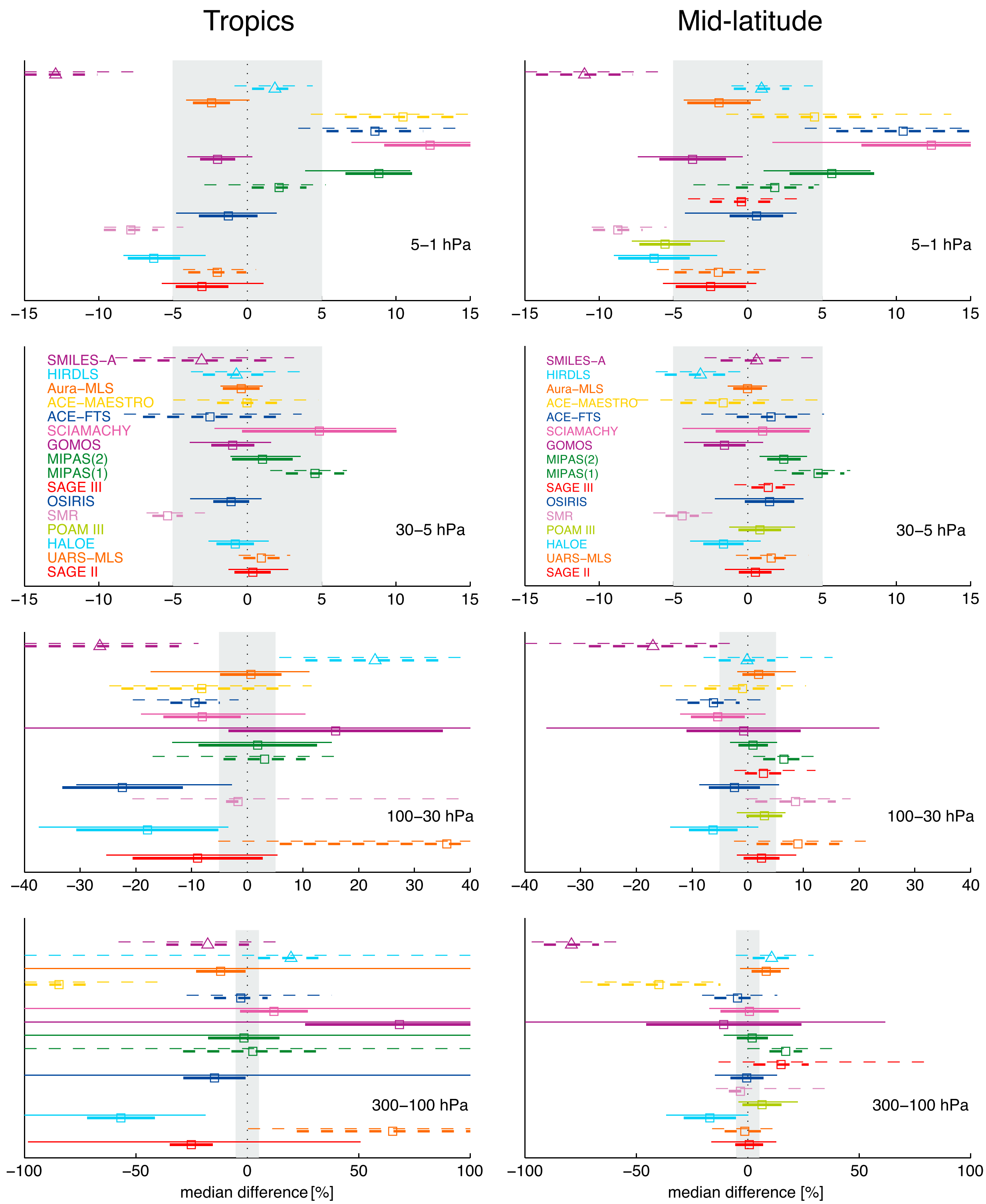

Figure 12. A summary of ozone differences for 1996–2010 is provided. Over a given latitude and altitude region, the median (squares), median absolute deviation (MAD, thick lines), and the standard deviation (thin lines) of the monthly mean relative differences between an individual instrument climatology and the MIM are calculated. Results are shown for the (left) tropics (20°S–20°N) and for the (right) midlatitudes (30°S–60°S and 30°N–60°N) and for four different altitude regions from the UT up to the US between 300 and 1 hPa.

apparent in the monthly mean data sets, have been discussed in past validation papers as well (see Table 2).

[54] LIMS and SAGE I provide the earliest ozone measurements and their climatologies agree very well in the MS with differences mostly within ±2.5% for all latitude bands. In the LS, differences are larger, up to ±20%.

[55] SAGE II provides the longest ozone data record with climatological ozone values in the tropics and midlatitudes that are in the middle of the measurement range given by the spread of all instrument climatologies. Exceptions are the tropical LS and UT, where SAGE II data show too low values compared to the other data sets, which is qualitatively consistent with a low bias in the SAGE II data in this region with respect to ozonesondes [*Wang et al*., 2002]. In the tropical and midlatitude MS, GOMOS and Aura-MLS climatologies show excellent agreement with the SAGE II climatology (differences below ±2.5%) while UARS-MLS, HALOE, OSIRIS, SAGE III, and MIPAS(2) agree very well with SAGE II with slightly larger differences (up to ±5%).

[56] HALOE and UARS-MLS observation periods overlap with SAGE II from 1991 to 2005 and 1999, respectively. The HALOE ozone climatology is in general low compared to the other data sets. The negative deviations of the HALOE climatology to the MIM are small in the MS and midlatitude LS (around −5%), larger but still in the climatological range in the US (−10%) and the tropical LS (−30%), and very large in the Antarctic LS in spring (−100%). The UARS-MLS climatology shows the opposite behavior compared to that of HALOE, with positive deviations from the MIM.

[57] POAM II, POAM III, and SAGE III mainly observe ozone at higher latitudes with a limited temporal coverage for some latitude bins that lead to larger biases in the annual means compared to the monthly means. The SAGE III climatology agrees very well with most other data sets, with only small differences from the MIM and with a narrow distribution. The POAM II climatology has a negative offset compared to other data sets which is particularly strong in the LS. The POAM III climatology shows a positive offset compared to the MIM, which is small in the MS (≤5%) and larger in the LS (~20%). Its sampling pattern allows POAM III to provide continuous solar occultation observations of the Antarctic ozone hole, where it reports more ozone than most other instruments (+20%) except during the peak of the ozone depletion at the end of SH winter, when it underestimates the ozone abundance (−50%).

[58] Among the newer data sets, OSIRIS, GOMOS, Aura-MLS, and HIRDLS climatologies in the MS/US are consistent and show only small deviations (e.g., average differences for the tropical MS of less than 1%). Aura-MLS performs exceptionally well in most regions, being in the middle of the range of all climatologies and providing a realistic characterization of ozone variability. While the other data sets also perform very well, they have some limitations. OSIRIS data in the SH are impacted by their limited sampling pattern, not allowing them to capture the seasonal cycle in the UTLS, and show somewhat larger differences from the MIM. The GOMOS climatology shows considerable disagreement with all other data sets below 30 hPa, including an unrealistic seasonal cycle and unrealistic spikes in the deseasonalized time series. The HIRDLS climatology agrees well with the MIM in most atmospheric regions except in the tropical LS, where it displays the strongest average deviation among all data sets of around +25%.

[59] SMR and SMILES provide the lowest climatological ozone values in the stratosphere. While SMILES agrees very well with the other instruments in the MS, differences of up to −20% are found in the LS and US. The SMR climatology agrees well with the other climatologies in the UTLS. However, above 30 hPa, it displays a negative offset which determines the lower boundary of the range of the climatological ozone data from all instruments. During Antarctic ozone hole events, SMR severely overestimates the ozone abundance by up to +100%.

[60] The ACE-FTS and ACE-MAESTRO climatologies agree well with those of the other instruments in the LS and MS. Both data sets have a positive offset in the US (+10%), and ACE-MAESTRO has a strong negative offset in the UT (−50 to −100%). In general, the differences of the two instruments' climatologies with respect to the MIM show very similar structures, which are opposite to that of the OSIRIS, Aura-MLS, and GOMOS climatologies. Largely as a result of their limited temporal sampling, their monthly zonal mean climatologies show larger differences at higher latitudes than most other instruments.

[61] The SCIAMACHY climatology shows in the early years a positive difference in the tropical stratosphere and midlatitude upper stratosphere of up to +20% which might be related to the vertical resolution of the ECMWF temperature data used in the SCIAMACHY retrieval and climatology construction. The differences are smaller after 2006, with maximum differences of up to +10%. SCIAMACHY provides a physically consistent data set but overestimates the QBO signal between 5 and 10 hPa and the Antarctic ozone during the time of the ozone hole (+50%).

[62] MIPAS measured with a different spectral and spatial resolution after 2005 and therefore provides two data products: MIPAS(1) and MIPAS(2). While the MIPAS(2) climatology shows mostly very small differences with respect to the MIM, the MIPAS(1) climatology has a positive offset up to 10% in the stratosphere and 20% in the troposphere. An exception to this classification is the US, where the MIPAS(1) climatology differences are smaller than ±5% and MIPAS(2) has a positive bias of around 10%. Due to the jump between the MIPAS data sets, analysis of time series from the complete MIPAS data requires a method which is immune against such discontinuities [e.g., *von Clarmann et al*., 2010].

7.4. Conclusions

[63] The evaluation of 18 ozone profile climatologies shows that our knowledge of the ozone atmospheric mean state is good in the tropical MS and in the midlatitude LS/MS. However, a large climatological spread in the tropical UTLS demonstrates the need for further evaluation activities in this region, including the use of existing in situ measurements from balloon or aircraft platforms and data sets from nadir sounders. Our findings show large interinstrument differences for monthly zonal mean ozone at high latitudes (compared to tropics and midlatitudes), which might be related to the different sampling patterns of the individual instruments. More detailed evaluations of high-latitude ozone (especially for ozone hole conditions)

will require the use of coincident measurement comparisons, polar vortex coordinates, and the incorporation of in situ measurements.

[64] Nearly all data sets show very good agreement in terms of interannual variability and are suitable for studies of climate variability. Note that some instruments show unrealistic spikes (month-to-month fluctuations) in some regions (e.g., GOMOS and ACE-MAESTRO). SAGE II has been used extensively in validation and long-term studies, and it is of interest to extend the time series through merging activities. As a result of their excellent agreement with SAGE II, the data sets from Aura-MLS, OSIRIS, GOMOS (only in the tropical and midlatitude MS/US), and MIPAS(2) (not above 10 hPa) are potential candidates for such merging activities without priori debiasing. The systematic ozone data set comparison presented here can serve as input for ongoing ozone-merging studies, such as (1) the SPARC/IO_3C/IGACO-O3/NDACC Initiative on Past changes in the Vertical Distribution of Ozone (SI2N), (2) the NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) Global OZone Chemistry And Related trace gas Data records for the Stratosphere (GOZCARDS) project, and (3) the European ozone Climate Change Initiative (ESA Ozone_cci), which aim to merge various sources into homogeneous data records suitable for trend studies.

[65] To improve future model-measurement comparison activities, evaluations of natural variability such as those presented here (seasonal cycle, interannual variability, downward propagating QBO signal) are recommended. Depending on the evaluation, individual instruments may need to be excluded from the comparison. Caution should be used when evaluating the seasonal cycle in the tropical LS, as this cycle is seen to vary in magnitude between the different instrumental climatologies, probably due to the different vertical resolutions of the instruments and the large vertical gradient of ozone in this region (J. L. Neu et al., The SPARC Data Initiative: Comparison of upper troposphere / lower stratosphere ozone climatologies from limb-viewing instruments and the nadir-viewing Tropospheric Emission Spectrometer (TES), submitted to *Journal of Geophysical Research*, 2013). More detailed comparisons with ozonesonde measurements are recommended.

[66] **Acknowledgments.** These climatological comparisons have resulted from the work of the SPARC Data Initiative team which has been ongoing since 2009. The authors thank the relevant instrument teams, the various space agencies (CSA, ESA, NASA, JAXA and other national space agencies), and organizations such as CEOS-ACC and IGACO. We particularly thank the ISSI in Bern for facilitating two successful team meetings in Bern as part of the ISSI International Team activity program and SPARC and WCRP for travel support. The work from S.T. was funded from the WGL project TransBrom and the European Union's Seventh Framework Programme FP7/2007–2013 under grant agreement n°226224 – SHIVA. M.I.H. thanks the CSA and ESA for supporting her work for the SPARC Data Initiative. Work at the Jet Propulsion Laboratory, California Institute of Technology, was performed under contract with the National Aeronautics and Space Administration. The ACE is a Canadian-led mission mainly supported by the CSA. Development of the ACE climatologies was supported by grants from the Canadian Foundation for Climate and Atmospheric Sciences and the CSA. Odin is a Swedish-led satellite project funded jointly by the Swedish National Space Board (SNSB), the CSA, the National Technology Agency of Finland (Tekes), the Centre National d'Etudes Spatiales (CNES) in France, and the third party mission programme of ESA. The work of E.K. was supported by the Academy of Finland through the project MIDAT (134325). The work of the University Bremen team was funded in part by the German Aerospace Agency (DLR) within the project SADOS (50EE1105) and by the State and University of Bremen. The work from Hampton University was partially funded under the National Oceanic and Atmospheric Administration's Educational Partnership Program Cooperative Remote Sensing Science and Technology Center (NOAA EPP CREST). We thank ECMWF for service and support and the German Weather Service (DWD) for providing access to the ECMWF data base.

References

Austin, J., and N. Butchart (2003), Coupled chemistry–climate model simulations for the period 1980 to 2020: Ozone depletion and the start of ozone recovery, *Q. J. R. Meteorol. Soc.*, *129*(595), 3225–3249, doi:10.1256/qj.02.203.

Baron, P., et al. (2011), The Level 2 research product algorithms for the Superconducting Submillimeter-Wave Limb-Emission Sounder (SMILES), *Atmos. Meas. Tech.*, *4*(10), 2105–2124, doi:10.5194/amt-4-2105-2011.

Brasseur, G., and S. Solomon (1984), *Aeronomy of the Middle Atmosphere*, Reidel Publishing Company, The Netherlands.

Chipperfield, M. P., L. J. Gray, J. S. Kinnersley, and J. Zawodny (1994), A two-dimensional model study of the QBO signal in SAGE II NO_2 and O_3, *Geophys. Res. Lett.*, *21*(7), 589–592, doi:10.1029/94GL00211.

Chu, W. P., M. P. McCormick, J. Lenoble, C. Brogniez, and P. Pruvost (1989), SAGE II inversion algorithm, *J. Geophys. Res.*, *94*(D6), 8339–8351, doi:10.1029/JD094iD06p08339.

Degenstein, D. A., A. E. Bourassa, C. Z. Roth, and E. J. Llewellyn (2009), Limb scatter ozone retrieval from 10 to 60 km using a multiplicative algebraic reconstruction technique, *Atmos. Chem. Phys.*, *9*(17), 6521–6529.

Dupuy, E., et al. (2009), Validation of ozone measurements from the Atmospheric Chemistry Experiment (ACE), *Atmos. Chem. Phys.*, *9*(2), 287–343.

Froidevaux, L., et al. (2008), Validation of Aura Microwave Limb Sounder stratospheric ozone measurements, *J. Geophys. Res.*, *113*, D15S20, doi:10.1029/2007JD008771.

Funke, B., and T. von Clarmann (2012), How to average logarithmic retrievals?, *Atmos. Meas. Tech.*, *5*(4), 831–841, doi:10.5194/amt-5-831-2012.

Gille, J., et al. (2008), High Resolution Dynamics Limb Sounder: Experiment overview, recovery, and validation of initial temperature data, *J. Geophys. Res.*, *113*, D16S43, doi:10.1029/2007JD008824.

Gillett, N. P., and D. W. J. Thompson (2003), Simulation of recent Southern Hemisphere climate change, *Science*, *302*, 273–275.

Grooß, J.-U., and J. M. Russell (2005), Technical note: A stratospheric climatology for O_3, H_2O and CH_4 derived from HALOE measurements, *Atmos. Chem. Phys.*, *5*(3), 2797–2807.

Jiang, Y. B., et al. (2007), Validation of Aura Microwave Limb Sounder Ozone by ozonesonde and lidar measurements, *J. Geophys. Res.*, *112*, D24S34, doi:10.1029/2007JD008776.

Jones, A., et al. (2009), Evolution of stratospheric ozone and water vapour time series studied with satellite measurements, *Atmos. Chem. Phys.*, *9*, 6055–6075.

Kyrölä, E., et al. (2010), GOMOS O_3, NO_2, and NO_3 observations in 2002–2008, *Atmos. Chem. Phys.*, *10*(16), 7723–7738, doi:10.5194/acp-10-7723-2010.

Ling, X.-D., and J. London (1986), The quasi-biennial oscillation of ozone in the tropical middle stratosphere: A one-dimensional model, *J. Atmos. Sci.*, (ISSN 0022-4928), *43*, 3122–3137, doi:10.1175/1520-0469.

Livesey, N. J., W. G. Read, L. Froidevaux, J. W. Waters, M. L. Santee, H. C. Pumphrey, D. L. Wu, Z. Shippony, and R. F. Jarnot (2003), The UARS Microwave Limb Sounder version 5 data set: Theory, characterization, and validation, *J. Geophys. Res.*, *108*(D13), 4378, doi:10.1029/2002JD002273.

Livesey, N. J., et al. (2008), Validation of Aura Microwave Limb Sounder O_3 and CO observations in the upper troposphere and lower stratosphere, *J. Geophys. Res.*, *113*, D15S02, doi:10.1029/2007JD008805.

Lumpe, J. D., et al. (1997), POAM II retrieval algorithm and error analysis, *J. Geophys. Res.*, *102*(D19), 23,593–23,614, doi:10.1029/97JD00906.

Lumpe, J. D., R. M. Bevilacqua, K. W. Hoppel, and C. E. Randall (2002), POAM III retrieval algorithm and error analysis, *J. Geophys. Res.*, *107*(D21), 4575, doi:10.1029/2002JD002137.

McCormick, M. P., J. M. Zawodny, R. E. Veiga, J. C. Larsen, and P. H. Wang (1989), An overview of SAGE I and II ozone measurements, *Planet. Space Sci.*, *37*(12), 1567–1586, doi:10.1016/0032-0633(89)90146-3.

McLinden, C. A., S. Tegtmeier, and V. Fioletov (2009), Technical note: A SAGE-corrected SBUV zonal-mean ozone data set, *Atmos. Chem. Phys.*, *9*(20), 7963–7972.

Mieruch, S., et al. (2012), Global and long-term comparison of SCIAMACHY limb ozone profiles with correlative satellite data (2002–2008), *Atmos. Meas. Tech.*, *5*(4), 771–788, doi:10.5194/amt-5-771-2012.

Molina, L. T., and M. J. Molina (1987), Production of chlorine oxide (Cl2O2) from the self-reaction of the chlorine oxide (ClO) radical, *J. Phys. Chem.*, *91*(2), 433–436, doi:10.1021/j100286a035.

Nardi, B., et al. (2008), Initial validation of ozone measurements from the High Resolution Dynamics Limb Sounder, *J. Geophys. Res.*, *113*, D16S36, doi:10.1029/2007JD008837.

Newman, P. A., E. R. Nash, S. R. Kawa, S. A. Montzka, and S. M. Schauffler (2006), When will the Antarctic ozone hole recover?, *Geophys. Res. Lett.*, *33*, L12814, doi:10.1029/2005GL025232.

Randall, C. E., et al. (2003), Validation of POAM III ozone: Comparisons with ozonesonde and satellite data, *J. Geophys. Res.*, *108*(D12), 4367, doi:10.1029/2002JD002944.

Randel, W. J., and F. Wu (2007), A stratospheric ozone profile data set for 1979–2005: Variability, trends, and comparisons with column ozone data, *J. Geophys. Res.*, *112*, D06313, doi:10.1029/2006JD007339.

Randel, W. J., M. Park, F. Wu, and N. Livesey (2007), A large annual cycle in ozone above the tropical tropopause linked to the Brewer–Dobson circulation, *J. Atmos. Sci.*, *64*(12), 4479–4488, doi:10.1175/2007JAS2409.1.

Ray, E. A., J. R. Holton, E. F. Fishbein, L. Froidevaux, and J. W. Waters (1994), The tropical semiannual oscillations in temperature and ozone as observed by the MLS, *J. Atmos. Sci.*, *51*(20), 3045–3052, doi:10.1175/1520-0469.

Remsberg, E., G. Lingenfelser, M. Natarajan, L. Gordley, B. T. Marshall, and E. Thompson (2007), On the quality of the Nimbus 7 LIMS version 6 ozone for studies of the middle atmosphere, *J. Quant. Spectrosc. Radiat. Transfer*, *105*(3), 492–518, doi:10.1016/j.jqsrt.2006.12.005.

Rousseeuw, P. J., and C. Croux (1993), Alternatives to the median absolute deviation, *J. Am. Stat. Assoc.*, *88*(424), 1273–1283, doi:10.2307/2291267.

Rusch, D. W., et al. (1997), Validation of POAM ozone measurements with coincident MLS, HALOE, and SAGE II observations, *J. Geophys. Res.*, *102*(D19), 23,615–23,627, doi:10.1029/97JD00458.

Santee, M. L., G. L. Manney, J. W. Waters, and N. J. Livesey (2003), Variations and climatology of ClO in the polar lower stratosphere from UARS Microwave Limb Sounder measurements, *J. Geophys. Res.*, *108*(D15), 4454. doi:10.1029/2002JD003335.

Schneider, N., F. Selsis, J. Urban, O. Lezeaux, J. D. La Noe, and P. Ricaud (2005), Seasonal and diurnal ozone variations: Observations and modeling, *J. Atmos. Chem.*, *50*(1), 25–47, doi:10.1007/s10874-005-1172-z.

Solomon, S. (1999), Stratospheric ozone depletion: A review of concepts and history, *Rev. Geophys.*, *37*(3), 275–316, doi:10.1029/1999RG900008.

Solomon, S., R. R. Garcia, F. S. Rowland, and D. J. Wuebbles (1986), On the depletion of Antarctic ozone, *Nature*, *321*(6072), 755–758, doi:10.1038/321755a0.

Son, S.-W., N. F. Tandon, L. M. Polvani, and D. W. Waugh (2009), Ozone hole and Southern Hemisphere climate change, *Geophys. Res. Lett.*, *36*, L15705, doi:10.1029/2009GL038671.

SPARC CCMVal (2010), SPARC report on the evaluation of chemistry-climate models, edited by V. Eyring, T. G. Shepherd, and D. W. Waugh, SPARC Report No. 5, WCRP-132, WMO/TD-No. 1526.

Steck, T., et al. (2007), Bias determination and precision validation of ozone profiles from MIPAS-Envisat retrieved with the IMK-IAA processor, *Atmos. Chem. Phys.*, *7*(13), 3639–3662.

Stiller, G. P., et al. (2012), Validation of MIPAS IMK/IAA temperature, water vapor, and ozone profiles with MOHAVE-2009 campaign measurements, *Atmos. Meas. Tech.*, *5*(2), 289–320, doi:10.5194/amt-5-289-2012.

Toohey, M., and T. von Clarmann (2013), Climatologies from satellite measurements: The impact of orbital sampling on the standard error of the mean, *Atmos. Meas. Tech.*, *6*, 937–948, doi:10.5194/amt-6-937-2013.

Toohey, M., et al. (2013), Characterizing sampling biases in the trace gas climatologies of the SPARC Data Initiative, *J. Geophys. Res. Atmos.*, *118*, doi:10.1002/jgrd.50874.

Urban, J., et al. (2005), Odin/SMR limb observations of stratospheric trace gases: Level 2 processing of ClO, N_2O, HNO_3, and O_3, *J. Geophys. Res.*, *110*, D14307, doi:10.1029/2004JD005741.

von Clarmann, T., et al. (2009), Retrieval of temperature, H_2O, O_3, HNO_3, CH_4, N_2O, $ClONO_2$ and ClO from MIPAS reduced resolution nominal mode limb emission measurements, *Atmos. Meas. Tech.*, *2*(1), 159–175.

von Clarmann, T., G. Stiller, U. Grabowski, E. Eckert, and J. Orphal (2010), Technical note: Trend estimation from irregularly sampled, correlated data, *Atmos. Chem. Phys.*, *10*(14), 6737–6747, doi:10.5194/acp-10-6737-2010.

Wang, H. J., D. M. Cunnold, and X. Bao (1996), A critical analysis of Stratospheric Aerosol and Gas Experiment ozone trends, *J. Geophys. Res.*, *101*(D7), 12,495–12,514, doi:10.1029/96JD00581.

Wang, H. J., D. M. Cunnold, L. W. Thomason, J. M. Zawodny, and G. E. Bodeker (2002), Assessment of SAGE version 6.1 ozone data quality, *J. Geophys. Res.*, *107*(D23), 4691, doi:10.1029/2002JD002418.

Wang, H.-J., D. M. Cunnold, C. Trepte, L. W. Thomason, and J. M. Zawodny (2006), SAGE III solar ozone measurements: Initial results, *Geophys. Res. Lett.*, *33*, L03805, doi:10.1029/2005GL025099.

Waugh, D. W., L. Oman, S. R. Kawa, R. S. Stolarski, S. Pawson, A. R. Douglass, P. A. Newman, and J. E. Nielsen (2009), Impacts of climate change on stratospheric ozone recovery, *Geophys. Res. Lett.*, *36*, L03805, doi:10.1029/2008GL036223.

Witte, J. C., M. R. Schoeberl, A. R. Douglass, and A. M. Thompson (2008), The quasi-biennial oscillation and annual variations in tropical ozone from SHADOZ and HALOE, *Atmos. Chem. Phys.*, *8*(14), 3929–3936.

WMO (2011), Scientific assessment of ozone depletion: 2010, Global Ozone Research and Monitoring Project-Report No. 52, Geneva, Switzerland.

Zawodny, J. M., and M. P. McCormick (1991), Stratospheric Aerosol and Gas Experiment II measurements of the quasi-biennial oscillations in ozone and nitrogen dioxide, *J. Geophys. Res.*, *96*(D5), 9371–9377, doi:10.1029/91JD00517.